Author:
Tseng, Hayley K

Title:
Cognitive Inhibition as a Core Component of Executive Functions
Exploring Intra- and Interindividual Differences

Cognitive Inhibition as a Core Component of Executive Functions: Exploring Intra- and Interindividual Differences

Hayley Tseng

School of Psychological Sciences

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy (PhD) in the Faculty of Life Sciences

September 2023

Word Count: 52, 945

Abstract

Cognitive inhibition is an essential executive function that we use in our everyday lives. Numerous factors have been claimed to influence this construct including video gaming, exercise and expertise in musical instruments. However, in this thesis, I focus on an understudied factor, the alignment of chronotype and testing time, and a heavily studied yet controversial factor, bilingualism. Throughout this thesis, with one exception, I present a series of experiments which have been conducted online. In the first empirical chapter, I examined a relatively novel Faces task which the authors have claimed to measure three cognitive processes, including two different forms of inhibition and task switching (Chapter 2). Based on this chapter's findings, I decided to use the Faces task in Chapters 3, 4 and 6. The next two chapters determined whether the alignment of time of testing and chronotype influences inhibition and task switching among the young adult (Chapter 3) and older adult (Chapter 4) population. Afterwards, I explored how conflict is resolved through a mouse tracking paradigm and by extension, whether this paradigm can be used for a variety of inhibition tasks (Chapter 5). For the final empirical chapter, I identified whether training inhibition in a verbal domain impacts inhibition in a non-verbal domain (i.e., far transfer effects). To achieve this, I investigated whether bilingualism, which can be seen as a form of cognitive training within the verbal domain, influences performance in non-verbal tasks which index inhibition (Chapter 6). The main findings of this thesis suggest that cognitive inhibition is not substantially impacted by synchrony effects nor by bilingualism. Furthermore, the findings imply that mouse tracking could be a promising tool to use to examine cognitive inhibition.

Acknowledgements

First and foremost, I must thank my supervisor, Markus Damian, for guiding me through this degree and for keeping me sane, especially during the COVID pandemic. Markus has been supportive and treated me with respect and kindness (even when I make one of my many lapses of judgment). I always felt heard and I appreciate the freedom he gave me when designing and writing the empirical studies discussed in this thesis. He always encouraged me to remain positive and resilient when things go awry. I could not have asked for a better supervisor and it was an absolute privilege to work with him.

I would also like to thank my lab group, Wenting Ye, Polly Barr, Will Chapman, Feng Ji and Lilith Zhu, it was always lovely to talk to you guys in our daily lab meetings. I am forever indebted to Will who helped collect the in-person data used in Experiment 2 of Chapter 5 (while I was stuck in Derbyshire).

I would like to express my gratitude to the Leverhulme Trust for funding my degree and thus, providing me with the extraordinary opportunity to study for a PhD. I owe a special thanks to Join Dementia Research (JDR) for providing me access to the recruitment pool of older adults. I could not have performed the study described in Chapter 4 without their help!

Finally, I would like to express how grateful I am to my family. They have always stood by me throughout both my life and PhD journey. Without their support, I would not have been in the position that I am in now. I would like to dedicate special kudos to my parents for putting up with my grumblings during the completion of this thesis!

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .... DATE:..........21/09/2023...............

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

A section of this chapter parallels parts of the Introduction section of:
Tseng, H., & Damian, M. F. (2023). Exploring synchrony effects in performance on tasks involving cognitive inhibition: An online study of young adults. *Chronobiology International*. https://doi.org/10.1080/07420528.2023.2256843

1.1 Introduction

We need cognitive inhibition to navigate everyday life. Consider the act of driving a car: to ensure you arrive at your destination safely, you need to dedicate your attention to the road and ignore distractions such as the social media notifications appearing on your phone. Previous research indicates that inefficient inhibition is associated with negative life outcomes such as gambling (e.g., Ioannidis et al., 2019; Smith, Mattick, Jamadar & Iredale, 2014) and obesity (e.g., Guerrieri, Nederkoorn & Jansen, 2012; Prickett, Brennan & Stolwyk, 2015; Favieri, Forte & Casagrande, 2019). Given this, it is important to explore the factors which influence how successful we are at ignoring irrelevant information or suppressing a prepotent action. In this thesis, I investigated how inhibition varies across different states within a person (i.e., intraindividual differences) and whether this process can be trained through bilingualism. The following paragraphs will first discuss the various taxonomies of inhibition, how this cognitive process is conventionally measured, and briefly review factors that may influence the efficiency of this process.

1.2 Defining inhibition

Inhibition can broadly be defined as the ability to resist attending to irrelevant information in order to achieve a goal. For example, a student may use inhibition to concentrate on studying for an examination rather than surfing the internet. Even though there are decades of research involving inhibition, a fundamental point of contention between researchers is the exact taxonomy of inhibition. Some researchers assert that this construct can be divided into three subprocesses (e.g., Hasher & Zacks, 1988) whereas others argue that inhibition can be split into eight subprocesses (Kornblum, 1994). For the sake of simplicity, I will describe two popular taxonomies of inhibition, including one proposed by Friedman and Miyake (2004) whose paper is cited 2,769 times at the time of writing, and the other generated by Hasher and Zacks (1988) whose paper is cited 4,927 times.

Friedman and Miyake (2004) proposed that inhibition can be represented by two processes. This proposal was based on a latent variable analysis to investigate how three potential subcomponents of inhibition relate to each other. These subcomponents were inhibition of prepotent responses (i.e., ability to avoid performing automatic responses), resistance to distractor information (i.e., ability to ignore irrelevant information) and resistance to proactive interference (i.e., the ability to ignore information which has recently become irrelevant). The authors observed a significant relationship between scores representing inhibition of prepotent responses and resistance to distractor interference and in turn, suggested a common underlying ability between them. By contrast, resistance to proactive interference was reported to be insignificantly related to the other two subcomponents. Considering these findings, the authors concluded that inhibition can be subdivided into two processes, response-distractor inhibition and resistance to proactive interference.

Hasher and Zacks' (1998) inhibitory model provides an alternative notion of inhibition. These authors compartmentalised inhibition according to its functions. Specifically, they proposed that inhibition consists of three functions: access, deletion and restraint. I would like to highlight that their definition of these functions heavily overlaps with the subcomponents defined by Friedman and Miyake (2004). The access function involves exclusively allocating attention to relevant information and hence, this function can be interpreted as parallel to the distractor interference defined by Friedman and Miyake. The deletion function helps an individual ignore irrelevant information which has somehow surpassed the access function or information which is no longer relevant. As you can see, this function is akin to the resistance to proactive interference process. Finally, Hasher and Zacks introduced the restraint function which involves suppressing dominant responses and hence, is similar to the inhibition of prepotent response process in Friedman and Miyake's framework. The authors explained that these three functions operate at different points of the information processing sequence. The access function is utilised when you first encounter the information while the deletion function is operated when the information is active in your mind and the restraint function is used when you are about to perform an action based on the information presented. Considering that the main focus for two of my chapters is on this model, my thesis will mostly follow the taxonomy described by Hasher and Zacks (1998).

1.3 How is inhibition measured?

Inhibition has been explored with tasks which ask participants to resolve some form of conflict where they have to make a response based on relevant information while attempting to ignore irrelevant ones. Conventionally, inhibition is determined by the difference in performance (usually, in terms of both error rate and reaction time) between trials which involve conflict (i.e., incongruent trials) and trials free from it (i.e., neutral trials) or trials which biases participants towards the correct response (i.e., congruent trials). Of course, I acknowledge that there are alternative methods of calculating inhibition. For instance, inhibition can be measured via the remote associates task where performance in leading distractors is compared with performance in misleading distractors. However, the main focus of my thesis will be on the conventional method used to capture inhibition. Three of the most popular conflict tasks used to measure this construct are Simon, Stroop and Flanker.

Simon Task (Simon & Rudell, 1967) In a typical trial of the Simon task, a participant would be provided an arbitrary mapping between a stimulus and a response (e.g., make a left response if you see a red triangle and make a right response if you see a blue square). Participants would be instructed to make a response based on the non-spatial feature of the stimulus while ignoring its location. The location would either be congruent (e.g., the red triangle appears on the left side of the screen) or incongruent (e.g., the red triangle appears on the right side of the screen) to the correct response. Relating back to Hasher and Zack's (1998) inhibition framework, successful performance in this task would mostly involve the use of the access and deletion functions because you are required to ignore an irrelevant piece of information. The ‘Simon effect' is calculated by the difference in performance between trials where the stimulus location and correct response are consistent (i.e., congruent trials) and trials where these two features are inconsistent (i.e., incongruent trials). For many studies, spatial information is the stimulus feature which participants have to ignore (e.g., Paap & Greenberg, 2013; Li et al., 2014; Paap et al. 2019, Xia et al. 2022). However, I would like to note there are other versions of the Simon task which deviate from this structure. For example, Kunde and Stöcker (2002) implemented a version of the Simon task where participants must generate their response through key press duration and the irrelevant information they had to ignore was how long the stimulus was presented.

Stroop Task (Stroop, 1935) Many versions of the Stroop task exist (see Macleod, 1991 for an overview). The most well-known version is the colour-word Stroop, where participants are asked to name the ink colour of the word. Other versions of this task, include the numerical Stroop (e.g., pick the number that has the larger font size), spatial Stroop (e.g., state arrow direction while ignoring the spatial location of the arrow) and emotional Stroop (e.g., state the ink colour of emotional and neutral words). A common feature of all these versions is that participants have to ignore the salient attribute of the target (e.g., in the colour word Stroop, the salient attribute would be the word name) and instead, focus on the less noticeable attribute (e.g., the ink colour). According to Hasher and Zack's (1998) model of inhibition, this task would involve the operation of all three functions of inhibition. The access and deletion functions would prevent the irrelevant information (e.g., word name) from being processed while the restraint function would help an individual resist performing the dominant response (e.g., saying the word name). Responses in this task can either be produced verbally (e.g. vocally name ink colour) or non-verbally (e.g. keypress). However, in this thesis, I will be using the non-verbal arrow version of the Stroop task (see Chapter 5 and Chapter 6).

Flanker Task (Eriksen & Eriksen, 1974) In a typical trial of the flanker task, participants are presented with a string of stimuli (commonly presented as letters or arrows) and are asked to identify the central stimulus. The stimuli surrounding the target (i.e., the flankers) would either elicit the same (i.e., congruent) or different (i.e., incongruent) response to the target. The 'flanker effect' is usually calculated by the difference in performance between the congruent and incongruent conditions. Here, the access and deletion functions would be required for this task.

A fairly consistent finding in papers which use these tasks is that participants produce slower response latencies and more errors for trials with conflict than those without it. As a result, there is a tendency for researchers to assume that these tasks are alike (Hommel, 2011) and in turn, to infer that these tasks can be used interchangeably. A consequence of this assumption is that there are studies which only used one of these tasks to measure inhibition in general (e.g., Costa et al., 2009; Schmidt et al., 2012; Martínez-Pérez et al., 2020). The issue with this is that, clearly, differences between these tasks exist. For instance, the tasks vary according to how conflict can successfully be resolved. To demonstrate, the Stroop task is more demanding on the ability to resist engaging in a prepotent response than the Simon task. To add to this, while the irrelevant information in the Simon and Stroop tasks is uni-modal, the irrelevant information in the flanker task is bi-modal. By this I mean that the relevant and irrelevant information dimensions in the flanker task are presented separately. As a consequence, participants could strategically perform some form of spatial filtering to ignore the distracting information in the flanker task but not in the Simon and Stroop tasks.

Previously, researchers have attempted to empirically demonstrate the differences between the tasks through delta plots. To clarify, delta plots help capture the time course of the experimental effects. This is achieved by plotting the effect of interest against overall response latencies. A positive slope implies that the experimental effect increases as a function of response latency while a negative slope suggests the reverse. While the flanker and Stroop effect have been reported to increase with response latencies, the Simon effect has been found to decrease (e.g., Pratte et al., 2010; Mittelstädt et al., 2022). There are various explanations as to why a negative sloped delta function was found for the Simon effect (for an overview see Pratte, 2021). For example, the negative slope may imply that activation of the irrelevant information occurs first and following this activation, inhibition slowly counters this congruency effect. Alternatively, this result may demonstrate that the motor activation evoked by the irrelevant information decays over time. Nevertheless, this pattern of results highlights that although these three tasks induce congruency effects, the technical characteristics of these effects vary from one another.

Identifying what exactly each conflict task measures becomes fundamental when comparing cognitive inhibition between different population groups. This is because you may expect performative differences between group X and group Y for tasks which measure one form of inhibition but not the other. By assuming that all the conflict tasks fundamentally measure the same form of inhibition, you may be ignoring the intricate details of the relationship between inhibition and the factor of interest. A good example of this relates to the bilingual advantage (more detail on this issue will be provided later in this chapter). Findings have been inconclusive as to whether bilinguals outperform monolinguals in tasks claimed to measure inhibition. However, there have been recent theories which indicate that bilingualism is only beneficial for specific forms of inhibition (see Chapter 6 for a more detailed overview) and if this is indeed true, then this would shed light on why findings have been mixed regarding this issue.

1.4 Can inhibition be reliably measured?

Recently, papers have challenged the notion that inhibition can be seen as a valid psychometric construct. A pioneering paper by Rey-Mermet et al. (2018) asked participants to perform a variety of tasks which have been claimed to measure inhibition, including Simon and Stroop. They observed low correlations between the accuracy costs elicited by the tasks and hence, concluded that these tasks do not measure the same construct. Instead, the costs evoked by these tasks represent task-specific effects rather than inhibition. Therefore, it is possible that inhibition is not a unitary construct as previously thought. Rey-Mermet et al.'s findings further highlight my point from the previous paragraph that researchers need to understand the nature of conflict present in a given inhibition task.

There are other potential reasons why Rey-Mermet et al. (2018) reported a lack of inter-task correlations. A possible reason relates to the 'reliability paradox'. This term was coined by Hedge, Powell and Sumner (2018) who state the fact that these conflict tasks are built to measure robust experimental effects. However, the features that are used to make these tasks experimentally robust would at the same time render the tasks unreliable for psychometric analysis. Put another way, in order to make an experimental effect robust, between-participant variability would have to be low. Indeed, for conflict tasks which are known to evoke robust experimental effects (e.g., Simon, Stroop and Flanker tasks), most people do perform better in the congruent than incongruent condition. Given that most conflict tasks are built to have low between-participant variability, the finding of low inter-task correlations is arguably inevitable.

1.5 Factors which influence inhibition and its controversies

In this section, I will introduce two controversial factors which have been claimed to influence cognitive inhibition and a potential confounding variable which may provide insight into why there are discrepancies in these research areas.

1.5.1 Cognitive aging

According to the Office for National Statistics, people in the U.K. are living longer. In 1991, 15.8% of the U.K. population were aged 65 years old or over. 25 years later, this percentage has increased to 18%. There are a plethora of societal and economic implications of an increasingly aging population. An implication of interest is that a longer living population means more people will experience the impact of cognitive aging (or sometimes termed 'cognitive decline'). Common complaints among older adults are regarding problems with memory and attention (Weaver Cargin et al., 2008). Indeed, it is widely agreed that abilities which rely on effortful processing, such as memory, are at their peak during young adulthood (Deary et al., 2009; Wilson et al., 2002; Zaninotto et al., 2018) but decline during middle and old adulthood. Of course, I should note that not all cognitive abilities decline with age. For example, abilities that are associated with knowledge have often been observed to be at their peak during old adulthood (e.g. Dumas, 2015; Salthouse, 2019). However, my thesis will mainly focus on the supposed negative effects of aging.

Recall how Hasher and Zacks' (1988) inhibitory deficit account provided a definition of inhibition which has been used in numerous studies. Not only did the authors provide a widely used definition of inhibition, but they also offered an account of why cognitive decline occurs. Their inhibitory deficit account argues that individual differences in cognitive functioning can be explained by differences in inhibition efficiency. According to this account, as people age, the efficiency of the access, deletion and restraint functions of inhibition decline.

Access Previous research implies that older adults are more likely to demonstrate deficits in the access function than young adults. This type of deficit would enable distractions to influence the processing of relevant information more easily. For example, during a shopping trip, a deficit would be demonstrated if you become more easily distracted by products that you did not intend to buy. In May (1999), older and younger adults had to perform a remote associates test (RAT). In a typical trial, three words were displayed and these three words were associated with a target word. The participant’s task was to identify this target word. For example, if a participant was presented with the words, 'sugar', 'candy' and 'walking', the correct answer would be 'cane'. Some of the trials involved distractors (i.e., a fourth word) which were either leading or misleading. The results indicated that older adults were significantly more hindered by the misleading distractors than younger adults. This demonstrates how distractors can disrupt task performance. Interestingly, the researchers also found that older adults benefited more from leading distractors than young adults. Not only does this finding further highlight the older adults' struggle to ignore distractors but it also dispels the perspective that this struggle immediately equates to negative consequences. Nevertheless, May's study illustrates age-related differences in the access function and this result has been reproduced in other studies including Kim, Hasher and Zacks (2007).

Deletion Prior studies have found that the efficiency of the deletion function declines with age. In Carpenter, Chae and Yoon (2020), younger and older adults were instructed to read a passage about a person going grocery shopping. In the distraction condition, participants were asked to ignore food-related words. By contrast, in the control condition, participants were asked to read the passage without any distracting information. After reading the passage, participants had to perform a creativity task where they were given three ingredients and had to produce as many recipes as possible. In comparison to the control condition, the distracting information used in the distraction condition appeared to facilitate creativity task performance more for older adults than younger adults. The larger carryover effect of the creativity task performance among older adults suggests that their deletion function struggled to remove the irrelevant information which bypassed the access function from the reading task.

Restraint Finally, there are studies which reflect age-related differences in the restraint function. Recall, the restraint function involves suppression of prepotent responses. For instance, you would require this function to win the game ‘red light, green light’. Prior studies indicate that older adults possess a deficit in this function when comparing their performance in the go/no-go task and Stroop task with young adults (e.g., Bedard et al., 2002). In this case, older adults were worse than young adults at overriding their urge to respond towards a cue in the no-go trials of the former task and resisting their prepotent tendency to read the word name in the latter task.

The inhibitory deficit theory claims that deficiencies in inhibition could explain why, for example, deficits in speech comprehension and working memory occur among older adults. Indeed, older adults have been observed to perform worse at the digit span task than young adults (Salthouse, 1994). The theory would argue that performance in this task could be influenced by inhibition because possessing fewer inhibitory resources means that an individual would have difficulty ignoring irrelevant information (e.g. previous digit span). As a consequence, the irrelevant information could hinder the ability to retrieve the relevant information (i.e. the current digit span). Of course, I acknowledge that working memory performance may not only be influenced by inhibition but also, for instance, the ability to generate and implement strategies to encode information and the ability to attend to relevant information. Nevertheless, an important argument made by the inhibitory deficit theory is that the efficiency of inhibition would, at least, influence the efficiency of other cognitive processes. Therefore, the theory assumes processes that require very little or no inhibitory resources remain stable with age.

Nevertheless, a potential argument against the inhibitory deficit account is that the results previously discussed reflect a deficiency in the activation of relevant information as opposed to the inhibition of irrelevant ones. For example, the worse performance in the Stroop task observed by Bedard et al. (2002) reflects older adults' struggle to activate information representing ink colour rather than their struggle to ignore information relating to word name. However, this alternate explanation does not convincingly address all the results. For example, the activation deficit account cannot explain why Carpenter et al. (2020) found, under the distraction condition, better creativity task performance evoked by the older adult than the young adult group. Moreover, prior studies report no significant differences in measures of activation between the two age groups. For example, May (1999) observed older adults performing Version 3 of the Extended Range Vocabulary Test (ERVT; Ekstrom, French, Herman & Dermen, 1976) better than young adults. For this test, participants were presented with a stimulus word along with five words. Their task was to identify which of the five words had the same meaning as the stimulus word. Here, it is assumed that this task requires the process of activation rather than inhibition. This is because the participants are required to access the meanings of all of the words and are not explicitly asked to ignore any irrelevant or withhold a dominant response. Given these findings, it appears that the cognitive decline that occurs in older age cannot entirely be explained by a decreased efficiency in activation.

More recently, however, questions have arisen of whether age-related differences in cognitive inhibition may have been 'exaggerated' (e.g., Verhaeghen, 2011, 2014). For instance, Rey-Mermet et al. (2018) empirically compared older and young adults' performance in a variety of inhibition tasks and reported inconsistent results. For the stop-signal and go/no-go tasks, young adults outperformed older adults, but this was not the case for the colour Stroop and flanker tasks. Adding to this, Rey-Mermet and Gade (2018) conducted a meta-analysis to explore whether older adults undergo a deficit in inhibition, and they concluded that in most tasks which are taken to measure inhibition, older adults do not show impaired performance compared to young adults. Researchers such as Verhaeghen attributed this inconsistency in findings to the fact that some of the studies reporting significant age-related differences have not controlled for processing speed. Indeed, a consistent finding within the cognitive aging literature is that older adults tend to perform tasks slower than young adults (e.g., Salthouse, 2000), and it could be argued that larger conflict scores found among older adults reflect the impact of slower processing speed rather than a deficit in inhibition. Overall, these recent findings imply that the age-related decline in inhibition may not be as robust as predicted by the inhibitory deficit theory and suggested by the early research.

1.5.2 Bilingualism

To clarify, in my thesis, I will refer to bilingualism as the ability to sufficiently use two languages. There is much research which investigates the mechanisms behind language comprehension and production among bilinguals (Grosjean & Li, 2013). To my knowledge, the general consensus among the literature is that the two languages in a bilingual's mind are always active (e.g., Goldrick et al., 2016; Nichols et al., 2021). One of the sources of supporting evidence for language coactivation are studies reporting a so-called 'cognate facilitation effect' in lexical decision tasks. A cognate is where a word in different languages share a similar form and meaning to each other. For example, 'winter' in English and 'winter' in Dutch share the same meaning and form. Bilinguals have been observed to respond faster to cognates than to control words that have different forms in the two languages (e.g., Schwartz et al., 2007) and words that share the same form but have different meanings (i.e., interlingual homographs; e.g., Lagrou et al., 2011; van Heuven & Dijkstra, 2010). A similar cognate facilitation effect emerges in tasks which require language production, such as in picture naming (e.g., Costa, Caramazza & Sebastian-Galles, 2000). These results imply that some form of co-activation of two languages takes place within a bilingual's mind.

The idea that language coactivation occurs when bilinguals undertake language production and comprehension is one of the core assumptions of the bilingual advantage hypothesis. The general rationale behind this hypothesis is that since the two languages seem to always be activated when a language is used, bilinguals will therefore have to regularly resolve competition between them (e.g., Bialystok, 2021; Green & Abutalebi, 2013). For example, an English-Spanish bilingual's intention of saying 'cat' may activate mental representations of both 'cat' and 'gato'. As a result, when engaging in a conversation about cats with an English speaker, the bilingual must resist accidentally saying 'gato'. To successfully resolve this type of competition, recruitment of domain-general processes including inhibition, working memory and attention may be required. Therefore, the positive bilingualism effect may arise due to these experiences of continuously managing two languages and the need for constant recruitment of supposedly domain-general processes including inhibition.

If bilingualism does incur benefits to inhibition, then this could be seen as an example of a 'far transfer' effect (i.e., applying a skill an individual learnt from one activity to another activity which is distinctly different). Indeed, there is evidence, albeit mixed, that these far transfer effects may occur for other activities such as playing a musical instrument (e.g., Abrahan et al., 2019; Rauscher & Hinton, 2011) and playing video games (e.g., Granic et al., 2014). If so, far transfer effects should arguably be expected to be especially strong for bilingualism. This is because activities like playing video games and playing a musical instrument would only occur for a limited number of hours whereas management of two languages can occur all day (Bialystok, 2017). The idea that bilingualism is associated with far transfer effects in cognitive inhibition leads to another major assumption of the bilingual advantage which is that linguistic and non-linguistic processing draw on the same cognitive network. Prior research does indicate that the brain regions used for language and non-verbal control overlap (e.g., Luk et al., 2011; Bialystok, Luk & Craik, 2012). However, as pointed out by de Bruin (2021), we would be relying on 'reverse inference' reasoning if we are to assume that this overlap indicates that linguistic and non-linguistic activities employ the same processes. This line of reasoning has been linked with the logical fallacy of affirmation of the consequence (if A then B, B therefore A).

A good deal of bilingual research has focussed on cognitive inhibition differences between young monolinguals and bilinguals. Initially, the idea that a cognitive benefit can be gained through having to manage two languages was seen as generally robust amongst researchers (Sanchez-Azanza et al., 2017; de Bruin, Treccani & Della Sala, 2015). Numerous studies have reported superior performance, in the form of either smaller congruency effects or faster average response latencies or both, by bilinguals in non-verbal conflict tasks including the Stroop (e.g., Hernández et al., 2010), Flanker (e.g., Costa et al., 2009) and Simon (e.g., Bialystok et al., 2005; Bialystok & DePape, 2009) tasks. However, a seminal study by Paap and Greenberg (2013) found equivalent performance between bilinguals and monolinguals in the Flanker, Simon and antisaccade tasks, and the authors concluded that there is a lack of 'coherent' evidence for the bilingual advantage in inhibition. Indeed, null findings in other studies suggest that the evidence for a bilingualism effect is inconsistent (e.g., Gathercole et al., 2014; Paap & Sawi, 2014; Paap, Johnson & Sawi, 2015; Kousaie et al., 2015).

The publication of the Paap and Greenberg (2013) article could be seen as a turning point in the bilingual advantage literature. In fact, Sanchez-Azanza et al. (2017) reviewed the publication trends in the literature of this field between 2005 and 2015. They found that, during the 2014-2015 period, there was an upsurge of published papers challenging the bilingual advantage. Additionally, when exploring the number of citations of papers published in 2014, they observed that studies reporting null findings were cited more often than those reporting positive findings. These trends reflect increasing interest in null findings. Notwithstanding, studies showing evidence in support of a bilingual advantage were still being published post-Paap and Greenberg (2013) including Bialystok et al. (2014), Woumans et al. (2015) and Schroeder et al. (2016). Nevertheless, the shift in publication trends reflects the fact that the bilingual advantage was not as robust as previously thought. At this present time, there is yet a definite answer as to whether bilingualism positively influences cognitive inhibition.

1.5.3 Synchrony effects: A potential confounding factor that needs to be controlled for in cognitive aging and bilingual advantage research?

To my knowledge, few studies performing between-group comparisons of the type outlined previously (i.e., comparing the performance of young vs. older adults, and of monolinguals vs. bilinguals) mention the time of day in which the participants performed the study. Sometimes, to control for time of day effects, researchers may instinctively test all participants at similar times. In fact, May, Hasher and Stoltzfus (1993) performed an informal survey and reported that researchers would most likely test both older adults and young adults between 12pm and 6pm. On the face of it, testing participants at similar times seems to control for time of day effects since participants would be tested under a similar environment. However, research within the circadian rhythm literature may suggest otherwise. To clarify, circadian rhythm is often described as an internal clock that regulates the sleep-wake cycle on a roughly 24-hr basis. Circadian rhythm manifests itself in many physiological parameters such as body temperature, but it also has consequences for a wide range of human behaviour including attention, memory, and executive function (see Schmidt, Collette, Cajochen, & Peigneux, 2007, for a comprehensive overview of circadian effects on cognition).

Circadian rhythm can vary from person to person and as a result, individual differences in preferred time of testing exist. Some people may identify themselves as a 'morning person' where they feel most alert in the morning and in turn, prefer to perform difficult physical and cognitive activities at this time. In contrast, others may consider themselves as an 'evening person' where they feel most alert in the evening and thus, prefer to engage in physically and cognitively demanding tasks at this time. Many individuals do not exhibit a strong morning or evening preference and are therefore categorised as 'neutral types'. The term 'chronotype' captures this preference towards performing daily activities in the morning, evening, or somewhere in between (Levandovski, Sasso, & Hidalgo, 2013). An individual's chronotype can easily and reliably be captured via psychometric tools such as Horne and Ostberg's (1976) Morningness-Eveningness questionnaire (MEQ). MEQ scores and physiological measures of circadian rhythm, including hormone secretion and body temperature significantly correlate with each other (Bailey & Heitkemper, 2001; Horne & Ostberg, 1977; Nebel et al., 1996). Prior studies have also reported significant age differences in chronotypes (Cajochen et al., 2006; May & Hasher, 1998; Roenneberg et al., 2007; Yoon, May & Hasher, 1999), with young adults more likely to identify themselves as an 'evening type' than a 'morning type' but the reverse for older adults. Hence, as individuals age, their chronotype tends to shift from eveningness to morningness.

As a result of an individual's chronotype, aspects of cognitive performance may oscillate when assessed at various points across the day. Indeed, chronotype along with time of testing have been reported to influence performance in a variety of cognitive tasks measuring attention, working memory and verbal memory (e.g., Barner, Schmid, & Diekelmann, 2019; Facer-Childs, Boiling & Balanos, 2018; Intons-Peterson et al., 1999; Lehmann, Marks, & Hanstock, 2013; Maylor & Badham, 2018; Schmidt et al., 2007; Yang, Hasher & Wilson, 2007). The notion of 'synchrony' captures this interaction between chronotype and testing time: when assessment time is aligned with the time of day favoured by an individual's chronotype, cognitive task performance is superior compared to when they are misaligned (May & Hasher, 1998). Researchers have argued that evidence for synchrony effects is most robust for tasks which require heavy use of cognitive inhibition (May, Hasher & Healey, 2023, May & Hasher, 1998). To my knowledge, there is not a clear explanation as to why this is the case. In other words, why are synchrony effects more influential in the efficiency of inhibition than other cognitive processes such as task switching and working memory? I speculate that perhaps the proposed robust evidence supporting synchrony effects for inhibition tasks demonstrates how cognitively demanding it is to suppress a dominant response or ignore distracting information.

Given the impact that synchrony effects can have on cognitive task performance along with individual differences in circadian rhythm, careful consideration of testing times may be needed when making group-wise comparisons. For example, as mentioned earlier, young adults tend to have a preference for eveningness while older adults usually possess a preference for morningness. Therefore, testing both age groups at the same time could potentially exaggerate or mask any group differences. To demonstrate, testing both age groups in the morning could mask any age-related differences in cognition since the testing time is advantageous to older adults but a hindrance to young adults. Conversely, the reverse could occur if the testing time was in the afternoon. Furthermore, to my knowledge, there is yet a piece of bilingual advantage research that controls for synchrony effects (Poarch & Krott, 2019) or reports when participants completed the cognitive tasks. As a consequence, it may be difficult to disentangle the influence of bilingualism from the impact of synchrony effects. For example, it would be hard to ascertain whether participant X outperformed participant Y on the flanker task because the former was bilingual and the latter was monolingual, or it is because one was tested at their optimal time while the other was not. Considering the potential implication of not controlling for synchrony effects, it is important to establish whether synchrony effects are a major confounding factor that needs to be addressed when making between-group comparisons.

1.6 Outline of thesis

The main aim of my thesis is to gain a better grasp on what inhibition is and what factors influence this executive function. Chapter 2 reports a series of pilot experiments which explore the capabilities of a so-far underused experimental task (the Faces task) created by Bialystok et al. (2006). The main motivation behind these experiments was to confirm that the Faces task can reliably measure two forms of cognitive inhibition, inhibitory control and response suppression, as well as task switching, as claimed by Bialystok et al. Chapters 3 and 4 report experiments which attempted to determine whether components of inhibition and task switching are impacted by time of testing along with chronotype. The former chapter involves a population of young adults while the latter involves a population of older adults. As well as providing insight into whether inhibition varies across the day, the results from these studies will be informative regarding whether controlling for synchrony effects is necessary when making group-wise comparisons regarding cognitive inhibition. Chapter 5 describes an investigation of how conflict is resolved through a relatively novel experimental paradigm (‘mouse tracking’). Specifically, I explored whether conflict is resolved in a continuous or staged manner and whether the mouse tracking paradigm can be applied to a variety of inhibition tasks. Chapter 6 investigates the controversy surrounding the bilingual advantage and whether acquiring a second language results in better inhibition. The last chapter provides a general discussion which synthesises the results described in Chapters 2 to 6.

Chapter 2: Exploring the Faces task as a tool to measure cognitive inhibition

A figure of this chapter parallels one of the figures used in:
Tseng, H., & Damian, M. F. (2023). Exploring synchrony effects in performance on tasks involving cognitive inhibition: An online study of young adults. *Chronobiology International.* https://doi.org/10.1080/07420528.2023.2256843

2.1 Chapter aims

In this chapter, I will describe a series of pilot experiments which determine whether the Faces task can be used to measure three different cognitive processes, namely task switching, inhibitory control and response suppression. A critical implication of this chapter is that if I find that this task can successfully capture these three subcomponents, then I can use it for my planned multi-sessional studies (reported in the following chapters).

2.2 Introduction

This chapter can be seen as a precursor to Chapter 3 where the main focus is to examine whether a candidate task is suitable for my synchrony effect studies (see Chapters 3 and 4). A planned aspect of these studies is for them to be multi-sessional. Therefore, a fundamental consideration that I need to make when designing this type of study relates to the task/tasks I should include. Considering the different subcomponents of inhibition, it would have been optimal to ask participants to complete a battery of conflict tasks. However, asking participants to complete multiple tasks at multiple times across the day may result in fatigue and encourage high attrition rates. Therefore, an alternative way to examine whether synchrony effects influence numerous subcomponents of inhibition is to find a task which can elicit a variety of indices representing different cognitive processes, for example, a task akin to the Attentional Networking Task (ANT) where researchers can gain three types of measures: alerting, orienting and conflict. In this chapter, I will focus on a modified antisaccade task which claims to measure three processes that are similar to or can be representative of the access, deletion and restraint functions described by Hasher and Zacks (1988).

In a conventional antisaccade task, participants would be asked to focus on a fixation cross located at the centre of the screen and a cue would be displayed to the left or right side of the cross. Participants would be tasked to produce an antisaccadic eye movement away from the cue. It is assumed that the high efficiency of inhibition is reflected by fast response latencies and low error rates in this task. More specifically, as you can see, participants are required to overcome their reflexes to look towards a salient cue. Given this requirement, successful performance in this task would involve the operation of the ‘restraint’ function. A fascinating version of the antisaccade task was introduced by Bialystok, Craik and Ryan (2006). Here, the authors combined aspects of the faces version of the Posner cueing paradigm (e.g., Friesen & Kingstone, 1998) with the antisaccade task. In their task, participants encountered a face displayed on a computer screen which was followed by a stimulus. This stimulus was presented on either the left or right side of the screen. They were instructed to perform an ipsilateral response towards the cue if the eyes on the face were green. If the eyes on the face were red, participants had to perform a contralateral response. As well as varying in colour, the eyes varied in gaze direction which either predicted where the cue would appear correctly (i.e., congruent), or incorrectly (i.e., incongruent). Alternatively, the eyes were looking straight ahead (i.e., neutral). An example of a typical congruent trial would be if participants were presented with green eyes looking left and the cue then appeared in the left box. In contrast, an example of an incongruent trial would be if the participants were presented with green eyes looking left but the cue appeared in the right box. Eye colour acted as a relevant cue while gaze direction acted as an irrelevant cue.

Bialystok et al. (2006) claimed that their Faces task can measure three different executive processes: response suppression, inhibitory control and task switching. Response suppression was determined by the difference in response latencies between the ipsilateral and contralateral trials (note that Bialystok et al. used the terms ‘prosaccade’ and ‘antisaccade’ for these two conditions). This process can be seen as akin to the 'restraint' function of the inhibitory deficit theory. Inhibitory control was defined by the difference in reaction time between congruent and incongruent trials. This process can be seen as similar to the ‘access’ and ‘deletion’ functions described by Hasher and Zacks (1988). Task switching was the difference in response latencies between mixed and blocked trials. Arguably, task switching requires the ability to ignore a task that was previously relevant and hence, involves the ‘deletion’ function of inhibition. Indeed, the authors tested this task on young and older adults and found that the intended cognitive processes can be captured using the keypress version of the task. A later study by Bialystok and Viswanathan (2009) clarified that this task could also be used to measure inhibitory control, response suppression and task switching among 8-year old children.

Considering the potential to measure indices roughly representing each function of inhibition (i.e., access, deletion and restraint), this task may be an ideal candidate for my planned within-participants synchrony effect study (see Chapters 3 and 4). However, to my knowledge, there are very few studies which have used Bialystok et al.'s (2006) Faces task.

Therefore, in this chapter, I will be describing a series of pilot experiments which explored whether Bialystok et al.'s (2006) task can capture indices representing inhibitory control, response suppression and task switching. Furthermore, to reduce the chances of overinterpreting my findings, I will be reporting results evoked by the Bayes Factor analyses as well as ones provided by the conventional analyses. To clarify, the Bayes Factor is the likelihood of the data supporting one hypothesis over the other. This contrasts with conventional statistics where the p-value indicates whether the null hypothesis can be rejected. Therefore, as you can see, Bayesian analyses will enable me to identify how strong the evidence is in supporting the null or alternate hypothesis. Considering the benefits of Bayesian analyses, I will be reporting the output of these analyses not only in this chapter but in all the empirical chapters of this thesis.

2.3 Experiment 1

2.3.1 Method

2.3.1.1 Participants

I recruited 38 participants who were Psychology students at the University of Bristol (Males = 6, Females = 32, mean age = 21.4 years old) who received course credit for their participation and self-reported to being monolingual and native British English speakers. Additionally, they reported to having normal/corrected-to-normal vision. In order to complete this experiment, participants required a laptop or desktop computer. This experiment received ethical approval (code: 12121997085) and all participants provided their informed consent.

2.3.1.2 Materials

I used Gorilla (https://www.gorilla.sc; Anwyl-Irvine et al., 2020) to create and run my experiment. Gorilla is an experiment builder tool which allows studies to be run online. The cartoon faces were created on the software Paint. Due to the nature of online studies, the screen size of the computer and general hardware used to run the experiment varied between participants.

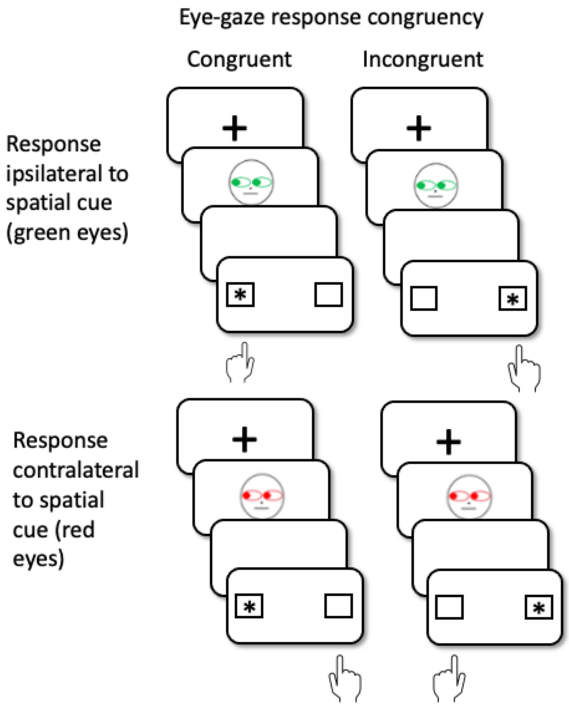
2.3.1.3 Design/Procedure

I replicated the trial design of the Faces task used by Bialystok et al. (2006). Eye colour was used as a relevant cue for participants to produce a response ipsilateral towards the spatial cue or a response contralateral to the spatial cue. Green eyes cued for the former (see Figure 2.1 for example) while red eyes cued for the latter (see Figure 2.1. for example). As you can see in Figure 2.1, eye gaze acted as an irrelevant cue and would either predict the position of the following cue correctly (i.e., congruent) or incorrectly (i.e., incongruent). I did not include the neutral trials as described by Bialystok et al. (2006), the reason being that eyes looking straight could arguably evoke social-cognitive consequences (e.g., Macrae et al., 2002). Response suppression was determined by the difference in performance between the ipsilateral and contralateral trials while inhibitory control was calculated by the difference in performance between the congruent and incongruent conditions. Finally, each trial was coded as either a task repeat (i.e., the previous and current trials involved the same response type) or a task switch (i.e., the previous and current trials involved different response types). As a result, I was able to calculate task switching by comparing performance between task repeat and task switch trials. As you can see, my definition of task switching differs from Bialystok et al. (2006) who compared performance between pure blocked and mixed blocked conditions. However, I felt using this definition would lengthen the task's running time which would not be ideal for my planned multi-sessional study.

Each trial began with a fixation cross located at the centre of the screen. After 250ms, a face was presented for 500ms which was then replaced with a blank screen for 200ms. The blank screen was finally replaced with two boxes with one of them containing an asterisk (i.e., spatial cue). In the ipsilateral condition (i.e., green eyes), participants were instructed to produce a response on the side at which the asterisk appears. For example, if the asterisk appeared in the left box, participants would have to press the left arrow button. In the contralateral condition (i.e., red eyes), participants would have to produce a response away from where the asterisk appears. For example, if the asterisk appeared in the left box, participants would have to press the right arrow key. If participants did not respond within 2,000ms after the boxes' appearance, the next trial would immediately start. In total, there were 200 trials which were presented in 5 blocks of 40 trials. Half of the trials were ipsilateral trials and the other half were contralateral ones. Additionally, half of the trials were congruent while the other half were incongruent. All trials were randomised for each block and the task took approximately seven minutes to complete.

Figure 2.1

Experiment 1. Example trials of all experimental conditions



2.3.2 Results

2.3.2.1 Pre-processing

Data was processed and analysed using the *afex* (Singmann, Bolker, Westfall & Aust, 2016), *bayestestR* and *emmeans* packages in *R* (R Core team, 2021). Participants were excluded for not responding to any of the trials (n = 11) and for producing an error rate higher than 25% (n = 8). Final analysis included data for 21 participants. For the response latency analysis, only response latencies for correct responses were analysed and reaction times greater than 1,000ms were excluded from the analysis. To compute the Bayes Factors, I used the default prior settings meaning the Bayes Factor is BF_{10} and its value represents how much the data relatively favours the alternate hypothesis (H_1) over the null hypothesis (H_0). Table 2.1. presents the guidelines used to interpret the Bayes Factors and I will be using this method of computing Bayes Factors throughout my thesis.

Table 2.1.
The guidelines used to interpret the Bayes Factors

Bayes Factor, BF_{10}			Interpretation
	>	100	Extreme evidence for H_1
30	-	100	Very strong evidence for H_1
10	-	30	Strong evidence for H_1
3	-	10	Moderate evidence for H_1
1		3	Anecdotal evidence for H_1
	1		No evidence
1/3	-	1	Anecdotal evidence for H_0
1/10	-	1/3	Moderate evidence for H_0
1/30	-	1/10	Strong evidence for H_0
1/100	-	1/30	Very strong evidence for H_0
	<	1/100	Extreme evidence for H_0

2.3.2.2 Final results

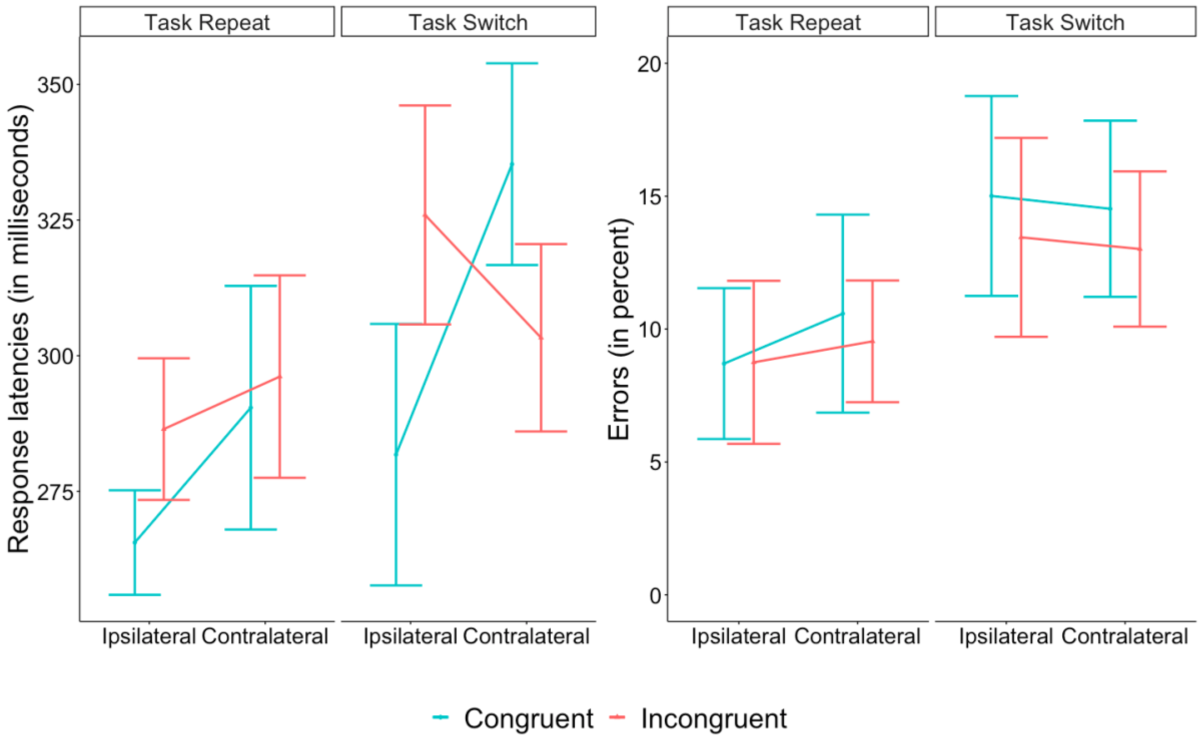
Table 2.2

Experiment 1. Analysis of Variance performed on response latencies and errors with within-participant factors, congruency (congruent vs. incongruent), response type (ipsilateral vs. contralateral) and task switch (task repeat vs. task swich). Significant results are bolded

Effect	df_1	df_2	MSE	F	η_p^2	p	BF_{10}
Errors							
Response type	1	20	81	0.10	.005	.756	0.18
Congruency	1	20	32	1.37	.064	.256	0.25
Response Type x Congruency	1	20	66	0.04	.002	.837	0.26
Task Switch	**1**	**20**	**60**	**14.92**	**.427**	**<.001**	**>100**
Task Switch x Response Type	1	20	31	1.08	.051	.310	0.30
Task Switch x Congruency	1	20	45	0.25	.012	.622	0.25
Task Switch x Response Type x Congruency	1	20	41	0.08	.004	.777	0.28
Response Latencies							
Response type	1	20	2,627	4.28	.176	.052	2.69
Congruency	**1**	**20**	**899**	**4.40**	**.180**	**.049**	**0.45**
Response Type x Congruency	**1**	**20**	**1,160**	**18.90**	**.486**	**<.001**	**57.95**
Task Switch	**1**	**20**	**3,015**	**10.10**	**.336**	**.005**	**>100**
Task Switch x Response Type	1	20	1,579	0.02	.001	.883	0.23
Task Switch x Congruency	1	20	1,099	0.50	.024	.488	0.26
Task Switch x Response Type x Congruency	**1**	**20**	**1,227**	**7.97**	**.285**	**.011**	**3.30**

Figure 2.2

Experiment 1. Response latencies (in milliseconds; left side) and errors rates (in percent; right side). Error bars reflect within-participants variability.



The output evoked from the ANOVA can be seen in Table 2.2. As you can see, the only main effect that emerged as significant was task switch with participants producing more errors in task switch (14.0%) than task repeat trials (9.4%) and Bayesian analysis lending 'very strong' evidence in favour of this finding. Like the results for the error analysis, the ANOVA for response latencies revealed a significant task switching effect with faster response latencies in task repeat trials (285ms) than task switch ones (312ms). However, unlike the output of the error analysis, I found a marginally significant response type effect with participants responding faster towards ipsilateral (290ms) than contralateral ones (306ms) but the Bayes factor of 2.69 suggests only 'anecdotal' evidence for this effect. Also, I observed a significant congruency effect of 10ms but the Bayes factor of 0.45 suggests 'anecdotal' evidence against this effect. To add to this, I found a significant interaction between response type and congruency with Bayesian statistics lending 'strong' evidence for this finding. Upon further investigation, the paired sample *t*-tests suggest that this interaction stems from the fact that the congruency effect occurs in the expected direction (congruent trials eliciting faster responses than incongruent trials) for ipsilateral trials, $t(20) = 4.56$, $p < .001$, but, although not significant, occurs in the reverse direction for contralateral trials, $t(20) = -1.91$, $p = .07$. The nature of this interaction can also be found in Figure 2.2.

Interestingly, I found a significant three-way interaction between response type, congruency and task switch. Figure 2.2 indicates that this finding is due to an interaction between congruency and response type emerging in task switch trials but not task repeat ones. The ANOVA supports this expectation with the interaction between congruency and response type being significant when analysing latencies with task switch trials only, $F(1, 20) = 19.03$, $MSE = 1{,}602$, $p < .001$, $BF_{10} > 100$, and not significant when analysing latencies with task repeat trials only, $F(1, 20) = 1.54$, $MSE = 785$, $p = .229$, $BF_{10} = 0.52$. When investigating the significant interaction involving task switch trials, paired sample *t*-tests indicate a congruency effect in the expected direction for ipsilateral trials, $t(20) = 3.29$, $p = .004$, but in the opposite direction for contralateral trials, $t(20) = -3.16$, $p = .005$.

2.3.3 Discussion

To recap, in this experiment, I used the Faces task to measure two types of inhibition, response suppression and inhibitory control, and task switching. Response suppression was calculated by the difference in task performance between ipsilateral and contralateral trials. In contrast, inhibitory control was defined as the difference in performance between the congruent and incongruent condition. Finally, task switching was identified by the difference in performance between task switch and task repeat trials. I expected to replicate Bialystok et al.’s (2006) keypress results for monolingual young adults. That is, I predicted to find a significant main effect of congruency (i.e., congruent vs. incongruent), which indexes inhibitory control, response type (i.e., contralateral vs. ipsilateral), which indexes response suppression and task switch (i.e., task switch vs. task repeat) which indexes task switching.

Firstly, I would like to highlight a noticeable difference in average response latencies between my experiment and Bialystok et al.'s (2006). While the average response latency was approximately 450ms in Bialystok et al. (2006), the average response latency in this experiment was approximately 295ms. Therefore, one could argue that it would be more difficult to identify the effects of interest in this experiment than in Bialystok et al. (2006). Despite the smaller response latencies, the results revealed significant main effects of congruency, response type and task switch in the expected direction, thus, suggesting that I was able to identify measures which supposedly index inhibitory control, response suppression and task switching. However, the Bayesian analysis only provided 'anecdotal' support for a significant response type effect and adding to this, the analysis offered 'anecdotal' evidence against a significant congruency effect. Considering the latter result and the very small congruency effect size of 10ms, I would argue that the Faces task was not able to sufficiently capture the process of inhibitory control, as claimed by Bialystok et al. (2006).

The significant interactions I observed in the ANOVA for response latencies may provide valuable insight into why the congruency effect size is very small. When further investigating the interaction between response type and congruency, I found a significant congruency effect in the predicted direction for ipsilateral trials only. Interestingly, however, for the contralateral trials, I found a reverse pattern, although non-significant, with faster response latencies in the incongruent than congruent condition. The three-way interaction between task switch, response type and congruency suggests that this reverse pattern result becomes significant when analysing task switch trials only. I can only speculate as to why a reverse congruency effect emerges for contralateral trials. For instance, the incongruent condition may have evoked faster responses because the irrelevant cue is indicating the correct response while the opposite is true for the congruent condition. Therefore, given these interactions, one could argue that the finding of near equivocal response latencies between the congruent and incongruent conditions is not surprising. As a consequence, a potentially better measure of inhibitory control is to calculate the performative differences between the congruent and incongruent conditions for ipsilateral trials only. Adding to this, I would further suggest that neutral trials should be added to my version of the Faces task and thus, response suppression should be calculated by the difference in performance between the two response types for neutral trials only. As I mentioned in the introduction section, having straight eyes as a neutral condition can be problematic and hence, an alternative method of introducing neutral trials is to have the eyes looking upwards. I implemented these changes in Experiment 2.

Another possible explanation for the very small congruency effect size relates to the Stimulus Onset Asynchronies (SOA) I implemented. SOA is the time between the onset of the cue and the onset of the target. In this experiment, I used an SOA of 700ms (as was the case in Bialystok et al., 2006) meaning that the target (the asterisk in one of the two boxes) was displayed 700ms after the cue (the cartoon face) was shown. It is common practice to use a variety of SOAs in saccade tasks, including button-press versions of it (e.g., Pashler, Carrier, & Hoffman, 1993). Previous studies have found that results vary according to the SOA used. To demonstrate, Friesen and Kingstone (1998) used a task similar to my Faces task, which primarily focussed on the congruent-incongruent performance, and found that the 'cueing effect' was present at 105ms, 300ms and 600ms SOAs. However, this effect did not arise for the 1005ms SOA which suggests that the appearance of the 'cueing effect' can be dependent on the SOA implemented. Through this rationale, it could be argued that the SOA I used was too long since participants were provided enough time to ignore eye gaze before the relevant cue was displayed. By using a shorter SOA, participants may not have enough time to ignore eye gaze and thus, response type and congruency effects may appear more strongly. I explored this possibility in Experiment 2.

2.4 Experiment 2 - Exploring the impact of SOAs

2.4.1 Method

2.4.1.1 Participants

43 participants (Males = 19, Females = 22, preferred not to say = 2, mean age = 23.7, SD = 3.96) were recruited through Prolific (https://prolific.co/) and they were paid for their participation. Prolific enables researchers to specify any pre-screening requirements for eligibility and when the study is published, Prolific will only allow eligible people to view the study. My pre-screening requirements included being under the age of 30 years old, only speaking English and no reports of diagnosed cognitive impairments. All participants reported to being monolingual British English speakers and having normal or corrected-to-normal vision. Additionally, only participants who had access to a laptop or computer could participate in this experiment. All participants provided their informed consent.

2.4.1.2 Materials

Like in Experiment 1, Gorilla was used to design and conduct the task. I used the same stimuli as the ones used in the previous experiment.

2.4.1.3 Procedure/Design

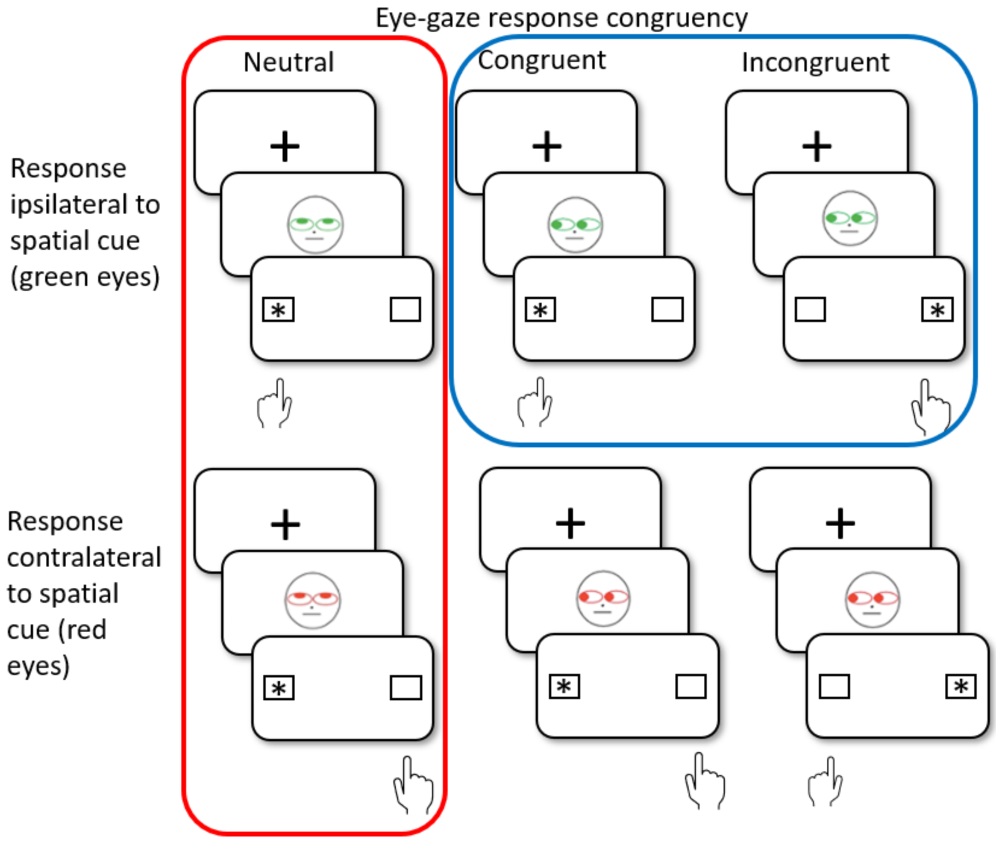
Broadly speaking, the experimental design used in this experiment mostly replicates the design implemented in the first experiment. However, here, I included practice trials, more experimental trials, and neutral trials where the eyes were looking upwards. Furthermore, I manipulated the stimulus onset asynchronies (SOA). By this I mean that the time interval between showing the cue (i.e., cartoon face) and the target (i.e., asterisk) varied between trials. In this case, I used three different SOAs: 300ms, 500ms and 700ms. In the SOA 300 condition, the cartoon face was presented for 100ms whereas in the SOA 500 condition, the face was presented for 300ms and in the SOA 700 condition, the face was presented for 500ms. For all these conditions, the following blank screen was displayed for 200ms. To reiterate, the independent variables in this experiment were congruency (congruent vs. incongruent vs. neutral), response type (ipsilateral response vs. contralateral response), task switch (task repeat vs. task switch) and SOA (300ms vs. 500ms vs. 700ms). The dependent variables were response latencies (in ms) and error rate (in percent).

As discussed in Experiment 1, analysing the outcome of this task via an omnibus ANOVA (as reported in Table 2.1 and Figure 2.2, and also reported in Bialystok et al., 2006) renders extraordinarily complex results which are difficult to interpret. The current study adds a manipulation of SOA, which multiplies the complexity of the design. As can be seen in Figure 2.3, for the current experiment as well as for all further studies in which I used the Faces task, I simplified the analytic protocol in order to gain a better handle on the target cognitive components of interest. Specifically, I changed how inhibitory control and response suppression were determined. Inhibitory control was calculated by the difference in performance between congruent and incongruent trials in the ipsilateral condition only. Response suppression was calculated by the difference in performance between the ipsilateral and contralateral conditions for neutral trials only. Like in Experiment 1, task switching was determined by the difference in performance between task repeat and task switch trials.

The instructions that the participants received in this experiment were the same as those used in the first experiment. Participants completed three blocks of 12 practice trials. They first completed a pure block of ipsilateral trials, then a pure block of contralateral trials and finally, a mixed block of ipsilateral and contralateral trials. After completing the practice trials, participants encountered 10 blocks of 36 experimental trials. This task took approximately 20 minutes to complete. Half of the trials required an ipsilateral response and the other half required a contralateral response. 120 trials were congruent, 120 were incongruent and 120 were neutral. Finally, there were equal numbers of trials which used an SOA of 300ms, 500ms and 700ms.

Figure 2.3

Experiment 2. Example of all experimental conditions. Response suppression was measured via comparison of the two conditions highlighted in red; inhibitory control was captured with a comparison of the two conditions highlighted in blue (see text for explanation).



2.4.2 Results

2.4.2.1 Pre-processing

Data was processed and analysed using the same R packages as used in Experiment 1. 14 participants were removed from the final analysis for having average error rates larger than 25%. 2 further participants were removed for incomplete data. In total, 27 participants were included in the final analysis.

Figure 2.4

Experiment 2. Response latencies. Panel A focuses on response suppression, panel B focuses on inhibitory control and panel C focusses on task switching. Each panel shows the three SOA conditions (300ms vs. 500ms vs. 700ms). Error bars reflect within-participants variability.

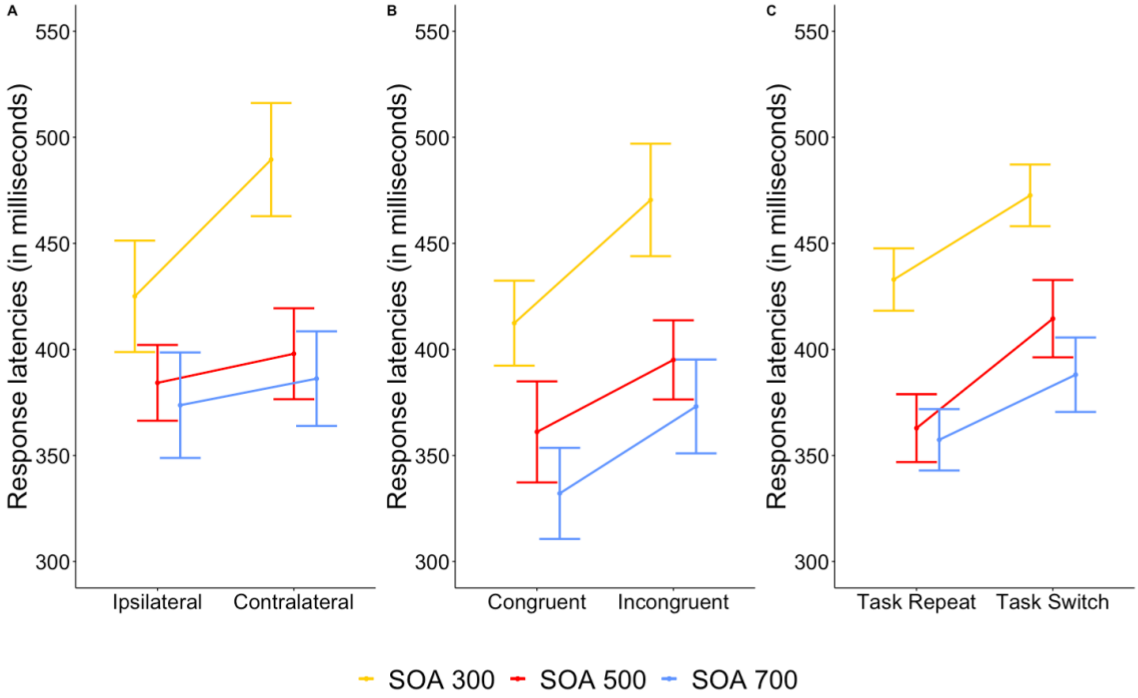
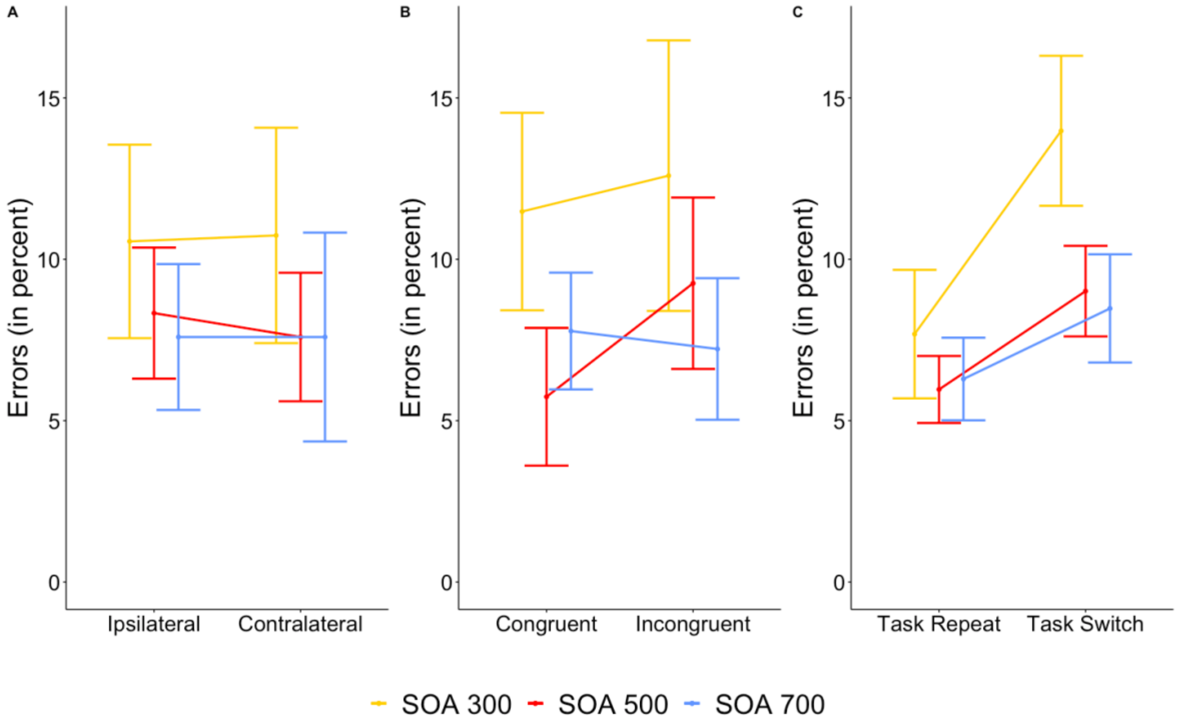


Figure 2.5

Experiment 2. Error rates. Panel A focuses on response suppression, panel B focuses on inhibitory control and panel C focusses on task switching. Each panel shows the three SOA conditions (300ms vs. 500ms vs. 700ms). Error bars reflect within-participants variability.



2.4.2.2 Response suppression analysis

A two-way ANOVA with within-participant factors response type (ipsilateral vs. contralateral) and SOA (300ms vs. 500ms vs. 700ms) revealed a significant response type effect, $F(1, 26) = 13.13$, $MSE = 2{,}820$, $p = .001$, $BF_{10} = 15.69$, with faster response latencies towards ipsilateral trials (394ms) than contralateral ones (425ms). Additionally, I found a significant SOA effect, $F(2, 52) = 21.44$, $MSE = 4{,}397$, $p < .001$, $BF_{10} > 100$, with response latencies being slower in the SOA 300 condition (457ms) than in the SOA 500 (391ms), $t(26) = 5.63$, $p < .001$, and SOA 700 (380ms) conditions, $t(26) = 5.73$, $p < .001$. However, response latencies did not significantly differ between the latter two conditions, $t(26) = 0.86$, $p = .671$.

Critically, there was also a significant interaction between these two main effects, $F(2, 52) = 4.00$, $MSE = 2{,}970$, $p = .024$, $BF_{10} = 1.47$. As can be seen in Figure 2.4, this interaction is likely to stem from the fact the response type effect appears to be significantly larger in the SOA 300 condition than in the other two conditions. My prediction appears to be met when performing paired sample t-tests with the response type effect being significant only in the SOA 300 condition, $t(26) = 3.47$, $p = .020$, and the Bayesian statistic lending 'strong' support for this finding ($BF_{10} = 16.83$). While the Bayesian analysis lends ‘strong’ evidence for the significant main effects, the analysis only lends 'anecdotal' evidence for the interaction ($BF_{10} = 1.45$). Regarding the parallel analysis with error rates, I found no significant main effects or interaction between the two effects (Fs < 3.02 and ps > .057) with Bayesian analysis lending ‘moderate’ support for the null hypothesis relating to the response type effect ($BF_{10} = 0.16$) and the interaction ($BF_{10} = 0.12$). Bayesian analysis revealed only ‘anecdotal’ evidence in favour of there being a significant SOA effect ($BF_{10} = 1.14$).

2.4.2.3 Inhibitory control analysis

An ANOVA with factors congruency (congruent vs. incongruent) and SOA (300ms vs. 500ms vs. 700ms) indicated significant main effects of congruency and SOA (Fs > 24.71 and ps < .001); the former effect suggesting faster response latencies towards congruent (369ms) than incongruent conditions (413ms) and the latter implying slowest responses in the SOA 300 condition (441ms, $t(26) = 5.97$, $p < .001$ when compared to SOA 500 and $t(26) = 7.33$, $p < .001$ when compared to SOA 700) and no differences occurring between the SOA 500 (378ms) and SOA 700 conditions (353ms), $t(26) = 2.42$, $p = .057$. The Bayes factor suggested ‘very strong’ evidence in favour of these significant main effects (BF_{10} > 100 for both effects). Finally, the interaction between these effects was not significant, $F(2, 52) = 0.70$, $MSE = 2{,}969$, $p = .501$, and the Bayes factor of 0.17 suggested ‘moderate’ evidence supporting this finding.

For the error analysis, the ANOVA revealed only a significant SOA effect, $F(2, 52) = 11.56$, $MSE = 32$, $p < .001$, with most errors produced in the SOA 300 condition (12.0%), compared to the SOA 500 condition (7.5%, $t(26) = 3.64$, $p = .003$, and compared to the SOA 700 condition (7.5%), $t(26) = 4.01$, $p = .001$) and no difference in errors emerging between the SOA 500 and 700 conditions, $t(26) = 0.00$, $p = 1.00$. Bayesian analysis lends ‘strong’ support for there being a significant SOA effect in errors ($BF_{10} = 32.35$).

2.4.2.4 Task switching analysis

An ANOVA suggested a significant task switch effect, $F(1, 26) = 24.31$, $MSE = 2{,}755$, $p < .001$, with faster response latencies in task repeat trials (384ms) than task switch ones (425ms), and an SOA effect, $F(2, 52) = 77.16$, $MSE = 1{,}256$, $p < .001$, with again slowest responses in the SOA 300 condition (453ms, $t(26) = 10.67$, $p <.001$ when compared with the SOA 500 condition and $t(26) = 10.39$, $p < .001$ when compared with the SOA 700 condition) and equivocal response latencies between the SOA 500 (389ms) and SOA 700 conditions (373ms), $t(26) = 2.40$, $p = .060$. The two main effects did not interact with each other, $F(2, 52) = 1.03$, $MSE = 1{,}449$, $p = .365$. The Bayesian analysis revealed ‘very strong’ evidence in favour of the significant main effects of task switch and SOA (BF_{10} > 100 for both effects). To add to this, this analysis provides ‘moderate’ evidence in favour of the null finding regarding the interaction between task switch and SOA ($BF_{10} = 0.23$).

The findings of the error analysis paralleled those of the response latency analysis. Significant main effects of task switch, $F(1, 26) = 17.99$, $MSE = 33$, $p < .001$, and SOA, $F(2, 52) = 16.09$, $MSE = 13$, $p < .001$, emerged; with the former effect implying more errors produced in the task switch (10.49%) than task repeat trials (6.65%), and the latter effect suggesting that the SOA 300 condition evoked the most errors (10.83%, $t(26) = 4.36$, $p < .001$ when compared to SOA 500 and $t(26) = 4.26$, $p < .001$ when compared to SOA 700) and the SOA 500 (7.49%) and SOA 700 (7.39%) conditions eliciting equivocal amounts of errors, $t(26) = 0.24$, $p = .968$. To add to this, I found a significant interaction between task switch and SOA, $F(2, 52) = 4.14$, $MSE = 15$, $p = .022$. As can be seen in Figure 2.5, this interaction is likely to stem from the fact that the task switching effect can be found in the SOA 300 and 500 conditions but not the SOA 700 condition. This appears to be confirmed by paired samples *t*-tests, for SOA 300: $t(26) = 3.85$, $p = .008$, for SOA 500: $t(26) = 3.40$, $p = .023$, for SOA 700: $t(26) = 1.94$, $p = .400$. Finally, the Bayesian statistics provided 'strong' evidence in favour of the alternate hypothesis for the two main effects ($BF_{10} > 100$) but only 'anecdotal' evidence for the interaction ($BF_{10} = 1.70$).

2.4.3 Discussion

The main aim of this experiment was to determine whether SOA impacts the size of the congruency, response type and task switching effects. To achieve this, I implemented three different types of SOAs, 300ms, 500ms and 700ms (the SOA I used in Experiment 1). I predicted that all the desired effects (i.e., congruency, response type and task switching) would emerge clearest at SOA 300 because with such a short SOA, participants have the least amount of time to resolve any conflict created by the manipulated variables. Indeed, a consistent finding across all the ANOVAs, in both error rates and response latencies, is the emergence of the SOA effect. Specifically, I found that the SOA 300 condition elicited higher error rates and longer response latencies compared to the SOA 500 and SOA 700 conditions. This implies that participants found trials which implement an SOA of 300ms most difficult to respond to.

My results indicated significant congruency and task switching effects in response latencies across the three different SOAs. The fact that the congruency effects occurred across the three different SOAs indicates that even at SOA 700, participants did not have enough time to resolve the conflict generated between gaze direction and cue location. The magnitude of the congruency effect was larger in this experiment (41ms) than in Experiment 1 (10ms). This discrepancy indicates that the likely reason why I was unable to observe substantial evidence for a congruency effect in Experiment 1 is because of how I calculated this effect. In contrast to the congruency effect findings, I observed a significant response type effect for SOA 300 only. This finding is particularly interesting as I was able to report this effect to be significant, albeit only marginally, in my previous experiment. Arguably, a conclusion that can be generated from this observation is that the response type effect may not be robust when using an SOA of 700ms.

From these findings, one could conclude that the simplified analytic protocol which was used to analyse the results (see Figure 2.3) is a good way to capture and isolate the three cognitive components of interest (inhibitory control, response suppression and task switching). Further, these subcomponents can be best captured by using a short SOA. Then again, arguably, an alternative explanation for my findings is that the SOA 300 condition seemed especially difficult when compared to trials of other SOAs. Therefore, the effects of interest (i.e., congruency, response type and task switching) may disappear once I remove the easier SOA 500 and 700 conditions. Given this argument, in the next experiment, I investigated whether the congruency, response type and task switching effects would still emerge if I only included trials with SOA 300.

2.5 Experiment 3 - SOA 300 only

2.5.1 Method

2.5.1.1 Participants

I recruited 40 University of Bristol students (Females = 18, Males = 21, mean age = 22.7 years old) and they received experimental credits for their participation. Again, participants self-reported themselves as monolingual and possessing normal/corrected-to-normal vision.

2.5.1.2 Materials

Like in Experiments 1 and 2, I used Gorilla to create and run my experiment. I used the same cartoon faces and cues as the ones used in the previous experiments.

2.5.1.3 Design/Procedure

I used the same design as Experiment 2 except I only included trials with SOA 300 and there were 360 experimental trials with equal proportions of congruency of congruent, incongruent and neutral trials. Also, there were equal numbers of ipsilateral and contralateral trials.

2.5.2 Results

2.5.2.1 Pre-processing

11 participants had to be excluded from the final analysis due to producing an error rate higher than 25%.

2.5.2.2 Final analysis

I performed separate one-way ANOVAs with the factor being the component of interest (congruent vs. incongruent for inhibitory control, ipsilateral vs. contralateral for response suppression and task repeat vs. task switch for task switching). For response latencies, I found the main effects of response type (27ms), congruency (43ms) and task switch (35ms) to be significant (all Fs > 10.67, ps < .003) and the Bayesian analysis indicated at least 'strong' evidence supporting these effects ($BF_{10} > 11.62$).

A parallel analysis of errors revealed no significant main effects of congruency or response type (Fs < .207 and ps > .106). However, the Bayesian analysis indicated only ‘anecdotal’ evidence in support of these null results ($BF_{10} > 0.50$). Only the task switch effect appeared to be significant, $F(1, 27) = 29.54$, $MSE = 4$, $p < .001$, with more errors produced in task switch (8.83%) than in task repeat (5.81%) trials. A Bayes factor > 100 implied ‘very strong’ evidence in favour of this significant finding.

2.6 General discussion

To summarise, the main aim of this chapter was to explore a candidate task, originally introduced by Bialystok et al. (2006), which could be used for my planned multi-sessional experiments reported in the upcoming chapters. The primary reason I chose to explore the Faces task was due to its potential to measure two forms of inhibition, response suppression and inhibitory control, as well as task switching, which arguably involves elements of inhibition. However, few studies have used the Faces task and hence, I found it imperative to identify whether the task can robustly measure the subcomponents of inhibition as claimed by Bialystok et al. (2006). Overall, capturing the indices representing inhibitory control and response suppression was more difficult than I expected. Recall, in Experiment 1, I was seemingly able to capture the process of response suppression but I did not observe convincing evidence to suggest that the task was able to capture inhibitory control. I speculated that this result could be explained by the fact that the SOA I used may be too long and the way in which I calculated these processes may have been flawed. Therefore, in Experiment 2, I addressed both issues. Interestingly, when looking at the results involving the SOA 700 condition only (the SOA I used in Experiment 1), I found a result reverse to Experiment 1, in that I was able to capture the process of inhibitory control but not response suppression. This perhaps implies that at SOA 700, the response type and congruency effects are not robust. Indices representing both processes, as well as task switching, were all identified at SOA 300. For that reason, in Experiment 3, I used a version of the Faces task where all the trials implemented this SOA to determine whether I could again capture these processes. Ultimately, the results suggest that these processes can indeed be captured when including SOA 300 trials only. Therefore, it does not appear to be the case that varying SOA is required for all the desired effects to emerge.

A noteworthy theme present in all three experiments is the fact that the Faces task was able to consistently capture the task switching effect regardless of the SOA used. This contrasts with the precise timings required to successfully observe the congruency and response type effects. Indeed, as mentioned in Chapter 1, different congruency effects (i.e., Stroop, Simon and Flanker) have been implied to possess different temporal profiles where the magnitude of some effects increase with time whereas others decrease (see Pratte et al., 2010, 2021, for discussion on this subject). Given this, one could argue that it is unsurprising that SOA influences various measures in contrasting ways. A possible conclusion that could be made is that participants are able to engage in inhibitory control and response suppression quicker than task switching. However, the reason as to why this could be the case is unclear and further research could investigate this issue. One definitive conclusion is that when using a short SOA, the Faces task is seemingly able to capture indices representing inhibitory control, response suppression and task switching. Although it should be noted that the corresponding effects of these processes are relatively small but robust. The effect sizes are comparatively smaller in my study than in Bialystok et al. (2006). To demonstrate, in Experiment 3, the response type, congruency and task switching effect sizes were 27ms, 43ms and 35ms while in Bialystok et al. (2006), the effect sizes were approximately 70ms (when comparing average latencies in Experiment 2 between ipsilateral and contralateral congruent trials in the mixed block condition for monolinguals only), 55ms (when comparing average latencies in

Experiment 2 between ipsilateral congruent and ipsilateral incongruent trials in the mixed block condition for monolinguals only) and 60ms (when comparing average latencies in Experiment 2 between the pure and mixed block condition for monolinguals only). A possible consequence of using a task which evokes small effect sizes is that intra- and inter-individual differences will be more difficult to capture. Given this, using the Faces task for my multi-sessional study is less than ideal in this regard.

The way in which I analysed the Faces task data in Experiment 3 is noticeably simpler than the analytical protocol used in Bialystok et al. (2006). In their study, they performed numerous ANOVAs including a five-way ANOVA with factors, language group (monolingual vs. bilingual), age (young adults vs. older adults), presentation condition (mixed vs. blocked), response type (ipsilateral vs. contralateral) and congruency (congruent vs. incongruent). Through this analysis, they claimed that they were able to capture effects representing inhibitory control, response suppression and task switching. However, I feel that this design is complicated and the effects found are difficult to interpret. For instance, as shown in Experiment 1, the nature of the congruency effect differed depending on the response type I was analysing. I found that the congruency effect was in the expected direction for ipsilateral trials regardless of the task switch condition. However, for contralateral trials, this effect was in the reverse direction when analysing task switch trials only and disappeared when analysing task repeat trials. Therefore, it is unclear what the congruency effect represents when using Bialystok et al.'s. (2006) definition of the effect. In contrast, I simplified the analytical protocol by altering how I calculated these processes. For example, I specified that inhibitory control was calculated by the difference in performance between the congruent and incongruent conditions for ipsilateral trials only. Response suppression was measured by the difference between the two response type trials for the neutral condition only. As a result, I was able to isolate the components of interest through simple one-way ANOVAs and what the effects represent is easier to interpret than in Bialystok et al. (2006). Having a simple analytical protocol to capture these components is desirable considering that I aim to use the Faces task for a multi-sessional study.

Finally, I would like to note that in my study, the Faces task was conducted online while Bialystok et al. (2006) was performed in person. As mentioned earlier, the online version of this task was seemingly able to capture the different components of cognition as claimed by Bialystok et al. (2006). This result is promising since it suggests that online testing can be a valid alternative to laboratory-based testing. However, it should be highlighted that I excluded a substantial number of participants due to high error rates (Experiment 1: 45%, Experiment 2: 37% and Experiment 3: 27.5%). Additionally, in Experiment 1, 11 participants did not complete any of the trials. This demonstrates a pitfall of online testing which is that the likelihood of encountering poor quality data is noticeably higher than in laboratory-based testing. As a result, the need to perform quality checks becomes especially important when performing online testing and to account for high attrition rates, one must recruit large sample sizes.

To conclude, given the results of Experiments 2 and 3, the Faces task does appear to be capable of capturing inhibitory control, response suppression and task switching when implementing an SOA of 300 and simplifying the analytical protocol, Therefore, I will use the version of the Faces task implemented in Experiment 3 to investigate whether inhibitory efficiency varies across the day (this experiment will be described in the next chapter).

Chapter 3: Synchrony effects among young adults

This chapter is based on:
Tseng, H., & Damian, M. F. (2023). Exploring synchrony effects in performance on tasks involving cognitive inhibition: An online study of young adults. *Chronobiology International.* https://doi.org/10.1080/07420528.2023.2256843

Note: The analytical protocol discussed in this chapter differs to the protocol described in the published paper. In the published paper, I calculated 'interference ratios' which were achieved by calculating the difference in response latencies between the critical and baseline conditions and dividing the result by the baseline condition. This was done for each participant, the component of interest and session time. To determine the presence of synchrony effects, I performed a series of regressions between the interference ratios and MEQ, separately for each component and session time. This contrasts with the analytical protocol reported in this chapter where I performed ANOVAs on raw response latencies with the factors, component of interest, session time and MEQ (treated as a continuous variable). Ultimately, the same theoretical conclusions were made from both of these analytical protocols.
Author contribution: All authors contributed to the conceptualisation, material preparation and data analysis of this study. HT wrote the first draft of the manuscript, both authors contributed to the final version.

3.1 Chapter aims

Previous studies suggest that it is beneficial to perform a cognitive task at a time which aligns with your chronotype. This type of finding has been described as the 'synchrony effect'. The inhibitory deficit theory argues that this effect is strongest for measures of inhibition and does not occur for measures of processing speed. So far, the evidence is not clear on whether this a robust effect among the young adult population. In this chapter, I describe an online study where I recruited a large sample of young adults who completed cognitive tasks in the morning, noon and afternoon. These tasks aimed to measure various forms of inhibition and processing speed. In turn, a critical theoretical implication of this study is that it will offer insight into whether synchrony effects need to be accounted for when making group-wise comparisons in inhibition and processing speed among young adults.

3.2 Introduction

The ability to inhibit irrelevant information and responses is clearly subject to individual differences (hence, allowing latent-variable analyses such as those featured in Friedman & Miyake, 2004). For example, previous studies indicate that young adults demonstrate stronger inhibition abilities than older adults (e.g., May, 1999; Bedard et al., 2002; Carpenter, Chae & Yoon, 2020). Furthermore, there have been findings which suggest that bilinguals exhibit stronger performances in tasks which measure inhibition than monolinguals (e.g., Bialystok et al., 2005; Bialystok, Craik & Luk, 2008; Cox et al., 2016) although it should be noted that this finding is still controversial (e.g., Paap, 2019 and see Chapter 6 for a more detailed discussion on this area). However, as mentioned in Chapter 1, to my knowledge, research in these areas tends not to control for any synchrony effects, the beneficial impact of aligning testing time with an individual's chronotype. A possible consequence of not addressing this issue is that group differences could be masked or exacerbated. For that reason, it is essential to identify whether researchers need to account for synchrony effects when conducting studies involving group-wise comparisons. The following paragraphs will provide an overview of evidence investigating this effect.

As briefly discussed in Chapter 1, research implies that synchrony effects can impact performance in numerous cognitive tasks. However, there have been claims that this effect does not emerge for all measures of cognitive performance and hence, the potential level of importance of controlling for this effect may depend on what aspect of cognition is being investigated. An interesting hypothesis advanced by May and Hasher (1998) is that synchrony effects are particularly pronounced in inhibitory abilities. However, it should be noted that the authors do not provide an explanation as to why synchrony effects are especially impactful on tasks measuring inhibition. In Chapter 1, I speculate that perhaps conflict tasks are more difficult than tasks which measure other cognitive processes. Nevertheless, in their pioneering study, they documented superior performance in the sentence completion task, which is presumed to index inhibition, among young adults tested at their preferred time of testing (i.e., the optimal group) than those tested at their non-preferred time of testing (i.e., the non-optimal group). Similar synchrony effects emerged in other tasks which require inhibition including the Sustained Attention to Response Task and memory tasks which involve ignoring specific stimuli (e.g., Lara, Madrid & Correa, 2014; May, 1999; Ngo & Hasher, 2017; Rothen & Meier, 2016). Furthermore, Lustig, Hasher, and Zacks (2007) suggested that while synchrony effects emerge in tasks that directly measure inhibition, they do not emerge in tasks that rely on automatic responses. For example, no synchrony effects were found in tasks that measured vocabulary, processing speed and general knowledge (Borella et al., 2010; Lara, Madrid & Correa, 2014; May, 1999; May & Hasher, 1998, 2017; Song & Stough, 2000), presumably because these activities do not directly involve cognitive inhibition. Therefore, one could conclude that researchers need to consider controlling for synchrony effects when making group-wise comparisons in inhibition but not in processing speed.

However, the consensus on whether a person's preferred time of testing needs to be considered when analysing their inhibition is still unclear. This is because as you can see in Table 3.1, which provides an overview of synchrony effect studies, the findings regarding the synchrony effect in inhibition among young adults are inconsistent. Significant synchrony effects among young adults have not always been found in tasks which are regarded as standard measures of inhibition (e.g., May & Hasher, 1998, Borella et al., 2010; Schmidt et al., 2012). For instance, as previously mentioned, May and Hasher (1998) reported a synchrony effect in the sentence completion task, but no such effect was found for young adults in more frequently used inhibition tasks including Stroop and stop-signal tasks (however, it should be noted that results were in the direction of worse performance in the non-optimal than optimal group). The reason as to why there is a discrepancy in findings between different measures of inhibition remains elusive.

Table 3.1
Summary of previous studies on synchrony effects in inhibition among younger individuals, by publication year

Reference	Design (between/ within Ss)	Sample (young)	Age range	Chronotype	Time window	Task	Non-behavioural measures included?	Behavioural results	Synchrony effect?
May & Hasher (1998), Exp. 1	between	N=48 (24 randomly assigned to AM and PM group)	17-21	screened to be of 'evening type' (MEQ score < 42)	AM: 8AM PM: 5PM	Sentence completion task, requiring suppression of no-longer-relevant information	no	Found negative priming for disconfirmed items at PM and positive priming at AM	✓
May & Hasher (1998), Exp. 2	between	N=36 (18 randomly assigned to AM and PM group)	17-21	screened to be of 'evening type' (MEQ score < 42)	AM: 8AM PM: 5PM	Stop-signal; Stroop colour naming (non-computerised)	no	Stop-signal task: stopping probability lower in AM than PM. Stopping time was lower in PM than AM but this did not reach significance Stroop: no significant in Stroop effect between AM and PM	?
Li et al. (1998), Exp. 1	between	N=32 (16 randomly assigned to AM and PM group)	17-20	screened to be of 'evening type' (MEQ score < 42)	AM: 8AM PM: 5PM	Reading aloud task with distracting words; recognition memory	no	No effect of time of day on reading performance	✗

Reference	**Design (between/ within Ss)**	**Sample (young)**	**Age range**	**Chronotype**	**Time window**	**Task**	**Non-behavioural measures included?**	**Behavioural results**	**Synchrony effect?**
Li et al. (1998), Exp. 2	between	N=32 (16 randomly assigned to AM and PM group)	18-21	screened to be of 'evening type' (MEQ score < 42)	AM: 8AM PM: 5PM	Reading aloud task with distracting words; recall memory	no	No effect of time of day on reading performance	✕
May (1999)	between	N=40 (20 randomly assigned to AM and PM group)	18-25	screened to be of 'evening type' (MEQ score < 42)	AM: 8AM PM: 5PM	Problem solving; measure ability to ignore leading and misleading distractors	no	Cost effect (cost of misleading distractors) was significant only for AM group. Benefit effect (benefit of leading distractors) was also only significant for AM group	✓
Intons-Peterson et al. (1999), Exp. 1	between	N=77 (randomly assigned to optimal and non-optimal testing groups)	18-25	implied everyone in this age group were 'evening types'	AM: before 10:30AM PM: at or after 3PM	Verbal false memory paradigm	no	No significant difference in recalled/recognised non-studied lures (indicating less efficient inhibition of non-relevant items) between the non-optimal and optimal groups	✕

Reference	**Design (between/ within Ss)**	**Sample (young)**	**Age range**	**Chronotype**	**Time window**	**Task**	**Non-behavioural measures included?**	**Behavioural results**	**Synchrony effect?**
West et al. (2002)	within	N=20 (10 randomly assigned to AM-PM or PM-AM testing order)	mean age = 24.4	predominantly 'neutral' or 'moderate morning'; not screened based on MEQ score	AM: 9AM PM: 5PM	Four-box task (spatial working memory task with distractors and n-back manipulation)	no	Numerically larger distractor effects in AM than PM, but performance difference not directly assessed statistically	?
Matchock & Mordkoff (2009)	within	N=80	18-28	no screening exclusion criteria. Participants were categorised as 'evening types' or 'morning'/'neutral types'	8AM, 12PM, 4PM, 8PM	Attention Network Test (ANT)	no	higher conflict scores at 12PM and 4PM regardless of chronotype group. No synchrony effect for orienting scores. However, synchrony effect did influence alerting scores for morning/neither type group only	?

Reference	Design (between/ within Ss)	Sample (young)	Age range	Chronotype	Time window	Task	Non-behavioural measures included?	Behavioural results	Synchrony effect?
Rowe, Hasher & Turcotte (2008)	between	N=56 (28 randomly assigned to AM and PM group)	18-39	Screened to be of 'evening type' (MEQ score < 42)	AM: 8AM or 9AM PM: 4PM or 5PM	Corsi Task (measure visual spatial working memory) - Participants completed a high interference (i.e., short list first) or low interference format (i.e., large list first)	no	Better performance from peak group than off-peak group for both formats	✓
Borella et al. (2010)	between	N=40 (20 randomly assigned to AM and PM group)	20-27	not measured/reported	AM: 8-11AM PM: 2-5PM	Stroop colour naming (single trial)	no	Stroop interference: (incongruent-control) similar between AM and PM Stroop facilitation (congruent-control): no difference between AM/PM	✕

Reference	**Design (between/ within Ss)**	**Sample (young)**	**Age range**	**Chronotype**	**Time window**	**Task**	**Non-behavioural measures included?**	**Behavioural results**	**Synchrony effect?**
Schmidt et al. (2012)	within	N=11	mean age= 25.6	screened for 'extreme eveningness' (MEQ <30)	AM: 9AM for external and 1.5 hours after waking for internal PM: 6PM for external and 10.5 hours after waking for internal	Stroop	no	Synchrony effect appeared for incongruent times appeared for external timing condition but then disappeared for internal timing condition	?

Reference	**Design (between/ within Ss)**	**Sample (young)**	**Age range**	**Chronotype**	**Time window**	**Task**	**Non-behavioural measures included?**	**Behavioural results**	**Synchrony effect?**
Knight & Mather (2013)	between	N=27 (randomly assigned to AM and PM group)	18-28	not measured/reported	AM: 8-10AM PM: 2-5PM	Attentional Network Task (ANT)	no	'Alerting' scores was numerically larger in AM than PM but this did not reach significance; no difference in 'orienting' or 'executive attention'	?
Lara, Madrid & Correa (2014)	within	N=27 (randomly assigned to complete AM or PM session first)	18-27	screened to be either a 'morning' or 'evening type'	AM: 8AM PM: 8:30PM	Sustained Attention to Response Task (SART)	no	Performance impaired over time in the non-optimal session but remained stable in the optimal session	✓

Reference	**Design (between/ within Ss)**	**Sample (young)**	**Age range**	**Chronotype**	**Time window**	**Task**	**Non-behavioural measures included?**	**Behavioural results**	**Synchrony effect?**
Rothen & Meier (2016)	between	N=160 (randomly assigned AM or PM session)	18-30	Participants asked whether they were a 'morning' or 'evening type'. Did not have category for 'neutral types'	AM: 6-10 AM PM:5-9PM	In the study phase, participants saw two superimposed picture, each coloured differently. They had to ignore pictures which were in certain colours. In the test phase, they performed the fragmented line task. Some of the pictures were also presented in the study phase.	no	Priming effects for ignored pictures were stronger in the non-optimal than optimal group	✓

Reference	Design (between/ within Ss)	Sample (young)	Age range	Chronotype	Time window	Task	Non-behavioural measures included?	Behavioural results	Synchrony effect?
Barclay & Myachykov (2017)	within	N=26	18-40	Chronotype was treated as a continuous variable (MEQ range: 36-63)	AM: 8AM PM: 2AM	Attentional Network Task (ANT)	no	No synchrony effects were found on ANT scores. There was some evidence of asynchrony effect with participants with a tendency towards eveningness exhibiting longer RT in PM than AM session and the reverse for those with a tendency towards morningness. At PM, participants with a tendency towards morningness was outperformed by those with a tendency towards eveningness in terms of RT and error rates on incongruent trials.	?

Reference	Design (between/ within Ss)	Sample (young)	Age range	Chronotype	Time window	Task	Non-behavioural measures included?	Behavioural results	Synchrony effect?
May & Hasher (2017)	between	N= 94 (randomly assigned to AM, noon and PM group)	17-21	Screened for 'neutral' (mean MEQ score of 52)	AM: 8AM Noon: 12PM PM: 5PM	Sentence completion; Stroop	no	Sentence completion: no significant priming effect were found at any time of the day. Stroop: Stroop effect did not vary across the day	✕
Ngo & Hasher (2017)	between	N=60 (randomly assigned AM or PM session)	17-27	screened for 'neutral' or 'evening type'	AM: 9AM - 12PM PM: 1PM-4PM	In phase 1, participants encountered cue words and had to either produce a word related or unrelated to the cue. In phase 2, they performed a lexical decision task	no	Compared to the control condition, participants were faster at responding to items in the unrelated condition at non-optimal time only. This suggests difficulty to suppress competitors	✓
Martínez-Pérez et al. (2020)	within	N=34 (randomly assigned to complete either AM or PM first)	Mean age = 21	Screened for 'morningness' and 'eveningness'	AM: 8AM PM:8:30PM	Flanker	no	Equivalent congruency effect between non-optimal and optimal testing times. Synchrony effect in terms of overall RT was found for 'evening types' only	✕

Reference	**Design (between/ within Ss)**	**Sample (young)**	**Age range**	**Chronotype**	**Time window**	**Task**	**Non-behavioural measures included?**	**Behavioural results**	**Synchrony effect?**
Rabi et al. (2022)	between	N=51 (randomly assigned to AM and PM group)	18-30	Screened for ‘eveningness’ (MEQ score < 46)	AM: 8-10:30AM PM: 2-5PM	Go-noGo; Flanker	EEG	Go-noGo: no significant time of day effects Flanker: no significant time of day effects found	✕

The lack of consensus in regard to synchrony effects is partly attributable to the scarcity of published behavioural studies which have investigated this issue, but many of the inconsistent results could also be attributed to small sample sizes (Richards et al., 2020). For example, the pioneering study by May and Hasher (1998) reported two experiments which used a between-participants design with 48 younger and 48 older participants, half of them tested in the morning and the other half tested in the evening. Barclay and Myachykova (2017) employed a within-participants design in which the performance of 26 younger individuals on the ANT was tracked across two times of the day. Despite the within-participant measurements, it is doubtful whether a sample of this size is substantial enough to reliably detect synchrony effects.

Another reason for the lack of consensus on synchrony effects in inhibition is that few studies have investigated this effect by including individuals with a 'neutral' chronotype. In fact, to my knowledge, there are only two studies that have explored this area (Barclay & Myachykova, 2017; May & Hasher, 2017). Most studies that explored synchrony effects only recruited participants who were either classified as a ‘morning type’ or an ‘evening type’, with participants who identified as a ‘neutral type’ excluded from the study or not invited to complete the cognitive tasks. For instance, Martínez-Pérez et al. (2020) asked 132 students to complete a MEQ and only participants who were either ‘morning types’ or ‘evening types’ were invited to the main study. A consequence of focusing on individuals who belong to an extreme group of a 'morning' or 'evening' chronotype is that I do not have a complete understanding of how the interplay of chronotype and time of testing impacts cognitive task performance. To my knowledge, May and Hasher (2017) is one of few studies which investigate synchrony effects among 'neutral types'. Young and older ‘neutral types’ were targeted in their study; participants completed tasks that measured a wide range of cognitive processes including inhibition, processing speed, memory and knowledge, with inhibition measured via a Stroop task. Participants were assigned to complete these tasks either in the morning (8am - 9am), at noon (12pm - 1pm) or in the afternoon (4pm - 5pm). The authors expected the noon group to perform better at the Stroop task than the other time groups, based on the assumption that for 'neutral types' the optimal time to perform cognitive tasks should be at noon. However, for the young participants, very similar Stroop effects were found in the three time groups. It is unclear why no synchrony effect was found in young 'neutral types', given that previous studies implied synchrony effects for young 'evening types' (e.g., May, 1999). One potential explanation, provided by the authors, for this disparity in results is that young 'neutral types' possess more cognitive flexibility than young 'evening types'. Then again, there is not a clear explanation of why this would be the case. Overall, May and Hasher's (2017) study provided mixed results regarding synchrony effects in 'neutral types'.

In summary, the notion of 'synchrony' in cognitive inhibition provides an intriguing theoretical concept but empirical support for it is at present mixed. Results for or against the psychological reality of synchrony effects are inconsistent, with some informative findings but also a good deal of null results. As I hoped to have highlighted above, it is challenging to a) identify adequate empirical tasks which could potentially capture synchrony effects and b) amass sufficient samples which would provide clear evidence for or against the synchrony concepts.

3.2.1 The present study

I tracked the performance of a large group, or at least when compared to the sample used by most of the previous studies, of young adults across three times of the day. A sample of this size (a total of over 100 participants were included in the critical analysis, following the exclusion of participants according to various criteria) was made possible by conducting the entire study online. To capture cognitive inhibition, I employed the Faces task (see Chapter 2 for a more detailed discussion of this task) whose purpose is to capture indices representing inhibitory control and response suppression. An additional component which was measured by the Faces task was task switching, i.e., whether the type of response (ipsi- vs. contralateral) was the same or different on two consecutive trials. Since Lustig et al. (2007) argue that performance which involves inhibition is affected by synchrony effects, task switching might also show such patterns. Evidence of whether this is the case is mixed. Bennett et al. (2008) reported a significant synchrony effect in performance in the Wisconsin Card Sorting Task (WSCT), but Yang et al. (2019) reported null findings on a cued match-to-sample task. Overall, however, research into the relationship between task switching and synchrony effects is scarce. Therefore, I sought to first establish whether a synchrony effect on task switching exists and if so, whether the magnitude of this effect is comparable to the effect on measures of inhibition.

As well as measuring inhibition and task switching, I included a measure of processing speed, which allowed me to assess the claim (Lustig et al., 2007) that synchrony effects emerge in activities which require cognitive inhibition, but that tasks which require automatic responses are unaffected. I captured processing speed with the Deary-Liewald task (Deary, Liewald, & Nissan, 2011) via the measurement of simple and choice reaction times. The former is measured by asking participants to respond to a single stimulus as quickly as possible while the latter is measured by asking participants to make an appropriate response to one of four stimuli. As for the Faces task, individuals' performance on the Deary-Liewald task was tracked at three time points of the day.

Throughout my study, I treated chronotype as a continuous variable (captured by MEQ score) rather than categorising individuals into chronotypes based on established cut-offs. This choice is not common in previous studies on this topic (see Table 3.1) and the standard approach in this field is to either not explicitly analyse chronotype at all (e.g., Knight & Mather, 2013) or to screen and select participants based on a particular chronotype (e.g., May & Hasher, 1998). In my study, I randomly sampled from the relevant population, and I felt that treating chronotype as a continuous variable would provide richer information than a reduction into standard chronotypes would, given that the cut-off points for different chronotypes originally suggested by Horne and Östberg (1976) are influenced by age (e.g., Taillard et al., 2004) and culture (e.g., Natale & Alzani, 2001; Pornpitakpan, 1998; see di Milia et al., 2013, for an overview of circadian typology). For this reason, I concur with Panjeh et al. (2021) that "…the idea that [morningness-eveningness] reflects a continuum is today regarded as a more appropriate way of characterising individuals than using coarsely categorised chronotype scores" (p. 234).

Ultimately, I aimed to explore whether a particular component of cognitive performance (response suppression, inhibitory control and task switching) can be predicted by a combination of MEQ and session time (morning, noon and afternoon). In this design, synchrony effects correspond to an interaction between MEQ and time of the day, and their effect on cognitive performance would emerge as an interaction between an inhibitory component, MEQ, and time of testing (in other words, a measure of inhibition or task switching is affected by the interplay between chronotype and time of the day). Separating inhibition into multiple components allows me to explore whether all, or perhaps only a subset of them, are affected by synchrony. Finally, the inclusion of the Deary-Liewald task allows me to investigate to what extent measures of inhibition dissociate from more general measures of processing speed.

3.3 Method

3.3.1 Participants

Initially, I recruited 332 young adult participants through a combination of Prolific (https://prolific.co/) and the participant pool of the University of Bristol. Participants recruited via Prolific received a monetary reward for their participation while the others received course credits. All participants were in the UK (GMT) time zone during testing. The study comprised four separate testing sessions (see below). 127 participants did not attend all the sessions and were therefore excluded from the analysis. Further exclusions were made before the final analysis and will be discussed in more detail in the Results section. All participants were self-reported as: monolingual, not colour-blind, not shift-workers, having not recently travelled abroad and not having sleep problems. All participants provided their informed consent and this study was approved by the University of Bristol Faculty of Life Sciences Research Ethics Committee (Approval code: 12121997085).

3.3.2 Materials

Morningness-eveningness questionnaire (MEQ). Chronotype was assessed through Horne and Ostberg's (1976) MEQ. This questionnaire consists of 19 multiple choice questions regarding preferred timings of activities and sleep-wake habits. An MEQ score, ranging from 16 and 86, is calculated by combining all the responses, with higher MEQ scores suggesting a greater morningness preference while lower MEQ scores indicating a greater eveningness preference. Conventionally, individuals are categorised into one of five chronotypes based on their score: 'definite morning' (70-86), 'moderate morning' (59-69), 'intermediate' (42-58), 'moderate evening' (31-41), and 'definite evening' (16-30). However, in my study, I treated MEQ and hence, chronotype as a continuous variable.

Background questionnaire. 14 questions in a background questionnaire elicited demographic information about the participants (i.e., sex, age, education, handedness and parents' education) and information about participants' level of involvement in activities including playing video games, using musical instruments and engaging in sports.

Session questionnaire. The session questionnaire was completed in the morning, noon and afternoon sessions, and consisted of 6 questions about participants' drug usage, alcohol intake, caffeine intake, hours of sleep, sleepiness (measured by the Visual Analogue Scale) and alertness (measured by the Stanford Sleepiness Score; Hoddes et al., 1973). To my knowledge, previous studies have not asked participants about these factors and as a result, arguably, it would be difficult to disentangle the impact of sleep, alcohol intake and drug use from any potential synchrony effects. Therefore, I attempted to control these confounding factors by excluding participants based on how they responded to these questions (see Pre-processing section for more detail on the exclusion criteria used).

Faces task. I mainly used the same design of the Faces task used in Experiment 3 of Chapter 2. The only notable difference is that I implemented fewer trials in this study (both practice and critical session versions) than in the one described in the previous chapter (360 trials). This was to reduce the likelihood of fatigue or boredom from performing the task multiple times. In the practice session (see below, 'Procedure'), the Faces task involved 32 practice trials and 72 experimental trials. In the three critical sessions, the Faces task consisted of 32 practice trials and 252 experimental trials. In both the practice and critical sessions, half of the trials involved ipsilateral trials and the other half involved contralateral trials. Furthermore, one third of the experimental trials were congruent, one third were incongruent and one third were neutral.

Deary-Liewald task. My measure of processing speed was adapted from Deary, Liewald and Nissan (2011). The task consists of two components: a simple reaction time (SRT) task and a choice reaction time (CRT) task. In a SRT trial, a box was displayed on the screen, and participants were instructed to press the spacebar key as quickly as possible whenever an 'X' appeared in the box. On each trial, one of six different wait times (i.e., time between the start of the trial and when the 'X' appeared) was randomly chosen: 400ms, 500ms, 700ms, 800ms, 1,000ms and 2,000ms. In a CRT trial, four horizontally aligned boxes were shown and an 'X' appeared randomly in one of the boxes, again following one of the six randomly chosen wait times. Participants pressed one of four designated response keys ('z' and 'x' keys, pressed with the index and middle finger of the left hand, and ',' and '.' keys pressed with the index and middle finger of the right hand) as quickly as possible. In both SRT and CRT trials, the 'X' was displayed for 1,000ms. All participants completed the SRT task first, followed by the

CRT task. In both the practice and critical sessions, this task consisted of 30 SRT trials and 30 CRT trials.

3.3.3 Procedure

Participants attended four consecutive online sessions. All participants completed a practice session on the day preceding the experimental phase. In this session, participants read the information sheet, provided informed consent, and completed the background questionnaire as well as the MEQ. Then, they completed practice runs of both the Faces task and the Deary-Liewald task. This enabled participants to familiarise themselves with the respective procedures, and minimised practice effects across the critical sessions which started the following day. Each participant completed three critical experimental sessions, conducted in the morning, noon, and afternoon. A third of participants completed the sessions in the order morning-noon-afternoon; a third completed the order noon-afternoon-morning (of next day for the morning session), and the remaining third completed the order afternoon-morning (next day)-noon (next day). Counterbalancing of the order was intended to further minimise the confounding of residual practice effects with session time. Morning sessions occurred between 8am and 9am (GMT Time), noon sessions occurred between 12pm and 1pm (GMT Time), and afternoon sessions occurred between 4pm and 5pm (GMT Time). Participants were informed to virtually attend these sessions through the messaging tools available at the University of Bristol and in Prolific. Additionally, Gorilla has a feature where researchers can track the exact time in which they perform tasks and questionnaires. Therefore, this feature allowed me to ensure that participants were attending the online sessions at the correct time. In each session, participants completed a session questionnaire, the Faces task and finally, the Deary-Liewald task. In the final session, participants were debriefed. Each session took approximately 25 minutes to complete.

3.4 Results

3.4.1 Pre-processing

Participants were excluded from both the final Faces and Deary-Liewald task analysis for the following reasons: reported to not being a young adult (n = 1), reported to having less than 5 hours of sleep before the first session and morning session (n = 15), reported to consuming recreation/prescription drugs which may impact cognitive performance (n = 10) and reported to consuming too much alcohol before at least one of the sessions (n = 3). Additionally, if the participant had error rates of more than 25% in one or both tasks, they were excluded from the final analysis of the task in which they performed poorly in. As a result, the number of eligible participants for the Faces task (n = 157) and Deary-Liewald task (n = 141) analysis differed slightly. Finally, I excluded random participants to ensure that for each task analysis, there were equal numbers of participants in each counterbalanced session order (e.g., whether the participants completed the morning, noon or afternoon session first). For the Faces task, I randomly excluded 12 participants from the noon first group and 1 participant from the afternoon group. For the Deary-Liewald task, I excluded 8 participants from the morning first group and 10 participants from the noon first group. The final analysis of the Faces task data consisted of 144 participants (mean age = 20.15 years old, Females = 102, Males = 42) while the analysis of the Deary-Liewald task data consisted of 123 participants (mean age = 20.15 years old, Females = 90, Males = 32, Prefer not to say = 1).

I used the software *R* (R Core team, 2021) with the packages *afex* (Singmann, Bolker, Westfall & Aust, 2016), *bayestestR* and *emmeans* for all statistical analyses. Response latencies over 2,000ms and under 150ms were removed for the response latencies aspect of the Faces task analysis (5.1% of trials excluded) and Deary-Liewald task analysis (0% excluded).

3.4.2 Faces task

3.4.2.1 Morningness-Eveningness Questionnaire demographics

Note that in my statistical analyses reported below, I did not use these categories but rather treated the MEQ score as a continuous variable. However, for my sample of 144 participants, the results were as follows: 'definite morning': n = 1 (0.7%); 'moderate morning': n = 15 (10.4%), 'intermediate': n = 91 (63.2%), 'moderate evening': n = 35 (24.3%), 'definite evening': n = 2 (1.4%). This profile of chronotypes among young adults converges with those found in earlier studies (e.g., May & Hasher, 1998, reported 0, 5, 58, 31, and 6% for the corresponding five chronotype types among their younger participants). Importantly, MEQ scores in my sample showed considerable variability (mean/median of 47, *SD* = 8.3, range = 27-70).

3.4.2.2 Response suppression

I compared performance on trials with ipsilateral (i.e., faces with green eyes) and contralateral responses (faces with red eyes) and this analysis was restricted to trials with neutral eye gaze. Latencies were analysed via a three-way Analysis of Variance (ANOVA), with response type (ipsilateral vs. contralateral), session time (morning; noon; afternoon) and MEQ score (entered as a continuous variable). Table 3.2 and Figure 3.1 show the results.

Table 3.2
Faces task. Analysis of Variance performed on response latencies, with response type (ipsilateral vs. contralateral), session time (morning vs. noon vs. afternoon) and Morningness-Eveningness Questionnaire (MEQ) score

Effect	df_1	df_2	MSE	F	η_p^2	p	BF_{10}
Response type	**1**	**142**	**2,025**	**40.97**	**.22**	**<.001**	**> 100**
Session time	**2**	**284**	**5,356**	**6.68**	**.05**	**.005**	**>100**
MEQ	1	142	56,697	0.39	<.01	.936	0.36
Response type × Session time	2	284	608	0.17	<.01	.961	0.03
MEQ × Response type	1	142	2,025	0.03	<.01	.961	0.16
MEQ × Session time	**2**	**284**	**5,356**	**4.14**	**.03**	**.040**	**33.82**
MEQ × Response type × Session time	2	284	608	0.04	<.01	.961	0.04

Note. P values were corrected via the Benjamini-Hochburg procedure (Cramer et al., 2016). Significant effects are bolded

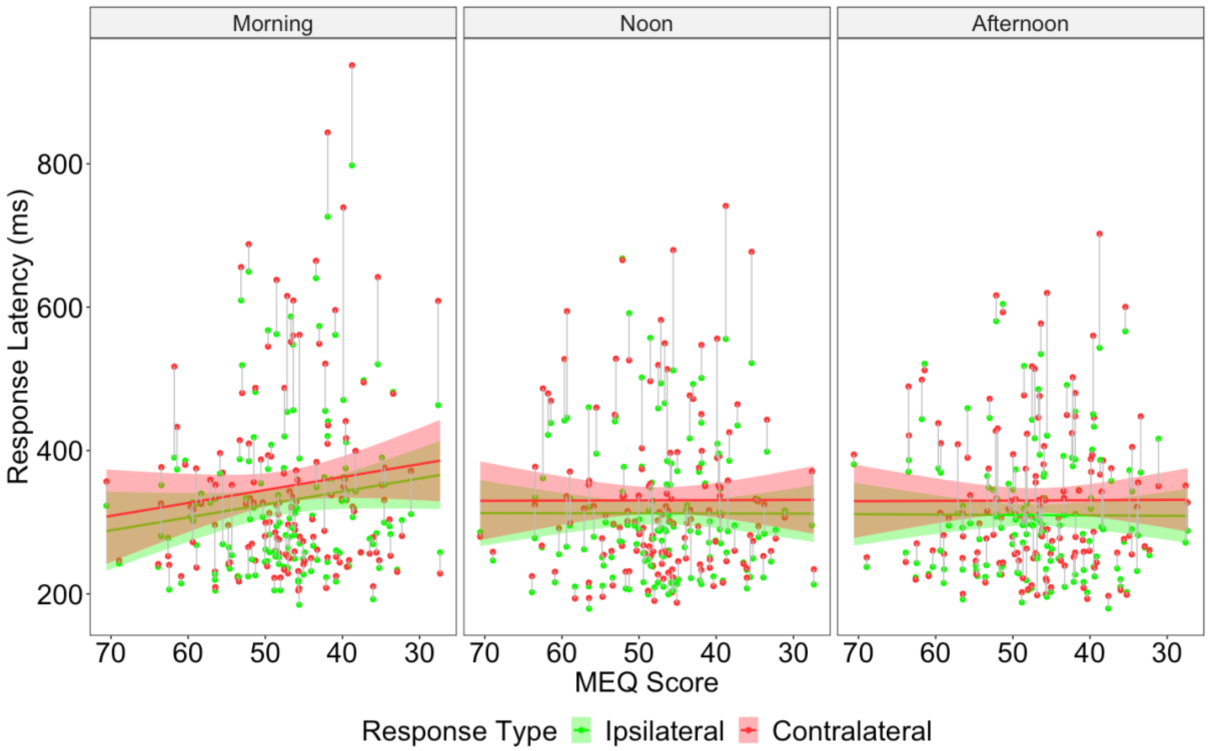
I found a highly significant effect of response type, with latencies 20ms faster in the ipsilateral than in the contralateral condition. Furthermore, a highly significant effect of session time was found (morning: 340ms; noon: 321ms; afternoon: 320ms). Follow-up tests revealed that morning latencies were significantly slower than latencies at noon, $t(142) = 3.06$, $p = .008$, and in the afternoon, $t(142) = 2.85$, $p = .014$, whereas the latter two did not differ significantly, $t(142) = 0.25$, $p = .966$. Finally, a significant interaction between session time and MEQ score was found. Figure 3.1 suggests a possible influence of MEQ score on latencies particularly in the morning session (as indicated by the positive slope of the fit line) but less so in the noon and afternoon sessions. However, follow-up tests showed no significant outcome for either session; Morning: $F(1, 142) = 2.25$, $p = .136$, Noon/Afternoon: Fs < .01, $p \geq .992$.[1] Critically, potential synchrony effects should emerge via a three-way

[1] Because the MEQ score is a continuous variable, the follow-up analysis amounts to a simple linear regression between the MEQ Score and latency. For the Morning session, the variance captured by this linear model was minimal, $R^2 = .02$. I additionally explored whether a non-linear model might characterise this relationship better,

interaction between response type, session time, and MEQ score: the size of the response type effect should be affected by the interplay (synchrony) between chronotype (as measured by MEQ) and session time. However, Figure 3.1 clearly suggests that the effects of response type were stable across the session times and not influenced by MEQ, and Table 3.2 shows that the three-way interaction was far from statistical significance. A Bayesian analysis conducted on the data showed a Bayes factor of 0.04 for the three-way interaction, rendering 'strong' support for the absence of synchrony effects in my results.

Figure 3.1

Faces task. Mean response latencies for ipsilateral (green) and contralateral (red) trials, dependent on Morningness-Eveningness Questionnaire score (MEQ; x-axis) and session time (from left to right: Morning, Noon, Afternoon session)



Note. X axis is reversed so that scores toward the left indicate a tendency toward an 'early' chronotype, and scores toward the right a 'late' chronotype. Dots represent participants and conditional means, with a grey line connecting the conditional means for a participant.

---

and I used "restricted cubic spline" analysis (Harrell, 2015) with either four or five knots, as suggested for samples of my size. However, these analyses did not result in a significantly improved fit relative to the simple linear regression model; the same was found for the other two components of inhibition (inhibitory control, and task switching) reported below. and I used "restricted cubic spline" analysis (Harrell, 2015) with either four or five knots, as suggested for samples of my size. However, these analyses did not result in a significantly improved fit relative to the simple linear regression model; the same was found for the other two components of inhibition (inhibitory control, and task switching) reported below.

A parallel analysis was conducted on error percentages. The only significant outcome was a main effect of response type, $F(1, 142) = 8.43$, $MSE = 26$, $\eta_p^2 = .056$, $p = .030$, $BF_{10} = 18.64$, with a higher error rate for ipsilateral (6.6%) than for contralateral (5.6%) responses. Hence, the possibility of a speed-accuracy trade-off regarding the effects of response type arises (faster RTs, but more errors, in the ipsilateral compared to the contralateral condition). To explore this possibility, I computed 'linear integrated speed-accuracy scores' (LISAS; Vandierendonck, 2017) in which latencies and errors were combined via the formula:

$$LISAS = RTj + \frac{\mathrm{SRT}}{\mathrm{SPE}} + PEj$$

where RT_j is a participant's mean RT in condition $_j$, PE_j is the corresponding error proportion, and S_{RT} and S_{PE} are the participant's overall RT and error standard deviations. The results from an ANOVA conducted on LISAS closely resembled the one reported in Table 3.2 for response latencies, with significant main effects of response type and session time and a significant interaction between MEQ and session time being found, Fs ≥ 3.57, ps $\leq .034$ and Bayes factor more than 8.25, but no significant three-way interaction between response type, session time, and MEQ score, $F = 0.19$, $p = .828$, $BF_{10} = 0.04$. Critically, the effect of response type on LISAS was still highly significant, $F(1, 142) = 19.78$, $MSE = 3{,}140$, $p < .001$, $BF_{10} > 100$, arguing against the notion of a speed-accuracy trade-off.

3.4.2.3 Inhibitory control

I compared performances on congruent (i.e., eye gaze corresponds to the correct response) and incongruent trials (i.e., eye gaze corresponds to the incorrect response). A three-way ANOVA with congruency (congruent vs. incongruent), session time (morning vs. noon vs. evening) and MEQ score (as a continuous variable) was used to analyse latencies. Table 3.3 and Figure 3.2 present the results.

Table 3.3

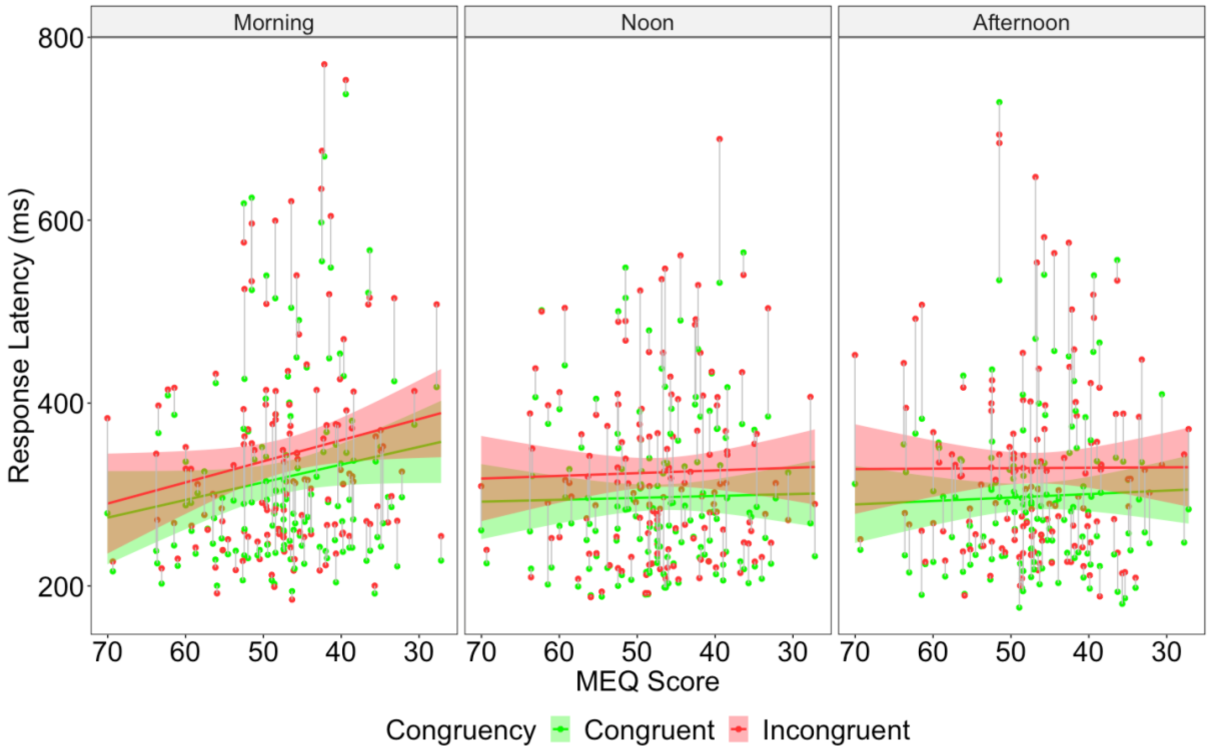
Faces task. Analysis of Variance performed on response latencies, with congruency (congruent vs. incongruent), session time (morning vs. noon vs. afternoon) and Morningness-Eveningness Questionnaire (MEQ) score

Effect	df_1	df_2	MSE	F	η_p^2	p	BF_{10}
Congruency	**1**	**142**	**1,526**	**106.68**	**.43**	**<.001**	**>100**
Session time	**2**	**284**	**4,937**	**7.23**	**.05**	**.003**	**>100**
MEQ	1	142	45,343	0.87	.01	.428	0.44
Congruency × Session time	2	284	687	1.31	.01	.428	0.04
MEQ × Congruency	1	142	1,526	<.01	<.01	.968	0.15
MEQ × Session time	**2**	**284**	**4,937**	**4.49**	**.03**	**.028**	**68.74**
MEQ × Congruency × Session time	2	284	687	1.01	.01	.428	0.04

Note. P values were corrected via the Benjamini-Hochburg procedure; significant effects are bolded.

Figure 3.2

Faces task. Mean response latencies for congruent (green) and incongruent (red) trials, dependent on Morningness-Eveningness Questionnaire score (MEQ; x-axis) and session time (from left to right: Morning, Noon, Afternoon session)



Note. X axis is reversed so that scores toward the left indicate a tendency toward an ‘early’ chronotype, and scores toward the right a ‘late’ chronotype. Dots represent participants and conditional means, with a grey line connecting the conditional means for a participant.

A significant main effect of congruency emerged: on average, participants responded 27ms faster in congruent trials than in incongruent trials. I also found a significant effect of session time (morning = 331ms, noon = 311ms, afternoon = 313ms), with latencies in morning sessions significantly slower than in noon sessions, $t(142) = 3.47$, $p = .002$, and afternoon sessions, $t(142) = 2.75$, $p = .018$, but no significant latency difference between noon and afternoon sessions, $t(142) = -0.54$, $p = .850$. Furthermore, a significant interaction between MEQ score and session time was found. Similar to Figure 3.1 showing the pattern for response suppression, Figure 3.2 indicates that MEQ scores influenced latencies only in the morning session, but a follow-up test suggests that this was not the case, $F(1, 142) = 3.57$, $p = .061$. Most importantly, as was the case for response suppression (see previous section) I found no evidence of a synchrony effect on inhibitory control: as shown in Table 3.3, the three-way interaction between session time, chronotype and congruency which would have indicated a potential synchrony effect was far from significance, and the impact of congruency appeared to be stable across the three sessions and the range of MEQ scores. A Bayesian analysis suggested 'strong' ($BF_{10} = 0.04$) evidence for the absence of the three-way interaction.

A parallel ANOVA conducted on error scores showed a significant main effect of congruency, $F(1, 142) = 80.47$, $MSE = 34$, $\eta_p^2 = .362$, $p < .001$, $BF_{10} > 100$, with participants making more errors in the incongruent condition (9.6%) than the congruent condition (6.0%). No other main effects or interactions were significant, $F\text{s} \leq 2.35$, $p\text{s} \geq .34$ and Bayes factors less than 0.21.

3.4.2.4 Task switching

Individual trials were coded with regard to whether the task (i.e., ipsilateral vs. contralateral responses relative to the spatial cue) on the previous trial was the same as on the current one. A three-way ANOVA with task switch (task switch vs. repeat), session time (morning vs. noon vs. afternoon) and MEQ score (as a continuous variable) is reported in Table 3.4.

Table 3.4

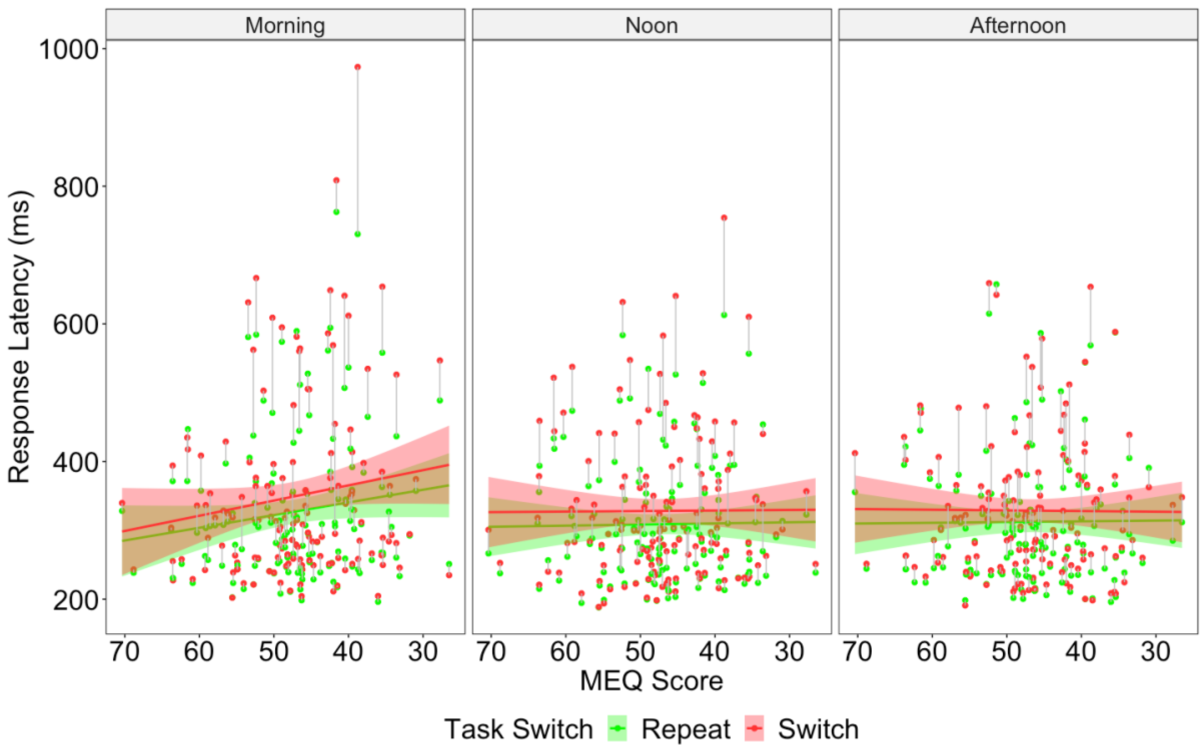
Faces task. Analysis of Variance performed on response latencies, with task switch (switch vs. repeat), session time (morning vs. noon vs. afternoon) and Morningness-Eveningness Questionnaire (MEQ) score

Effect	df_1	df_2	MSE	F	η_p^2	p	BF_{10}
Task switch	**1**	**142**	**1,117**	**72.25**	**.34**	**<.001**	**>100**
Session time	**2**	**284**	**4,802**	**7.27**	**.05**	**.003**	**>100**
MEQ	1	142	54,020	0.57	<.01	.528	0.41
Task switch × Session time	2	284	301	2.02	.01	.235	0.03
MEQ × Task switch	1	142	1,117	0.01	<.01	.920	0.15
MEQ × Session time	**2**	**284**	**4,802**	**5.10**	**.04**	**.016**	**>100**
MEQ × Task switch × Session time	2	284	301	1.43	.01	.337	0.04

Note. P values were corrected via the Benjamini-Hochburg procedure; significant effects are bolded.

Figure 3.3

Faces task. Mean response latencies for task repeat (green) and task switch (red) trials, dependent on Morningness-Eveningness Questionnaire score (MEQ; x-axis) and session time (from left to right: Morning, Noon, Afternoon session)



Note. X axis is reversed so that scores toward the left indicate a tendency toward an 'early' chronotype, and scores toward the right a 'late' chronotype. Dots represent participants and conditional means, with a grey line connecting the conditional means for a participant.

I found a significant main effect of task switch, with latencies 20ms longer on switch than on repeat trials, and a significant effect of session time, with slower latencies in the morning session (339ms) than in the noon (319ms), $t(142) = 3.43$, $p = .002$, and afternoon sessions (321ms), $t(142) = 2.75$, $p = .018$, but no significant difference between noon and afternoon sessions, $t(142) = 0.41$, $p = .913$. Furthermore, a significant interaction between MEQ score and session time emerged. Figure 3.3 implies that this interaction stems from the influence of MEQ score in the morning session. However, follow-up tests suggested that MEQ score did not significantly influence latencies in the morning session, $F(1, 142) = 2.89$, $p = .09$. Most importantly, the three-way interaction between session time, chronotype and task switch which could indicate a potential synchrony effect was far from significance, and Figure 3.3 suggests that the impact of task switch is stable across the three sessions and the range of MEQ scores. Bayesian analysis indicated 'strong' ($BF_{10} = 0.04$) evidence for the absence of the three-way interaction.

A parallel ANOVA conducted on error scores showed only a significant main effect of task switch, $F(1, 142) = 17.33$, $MSE = 87$, $\eta p^2 = .38$, $p < .001$, $BF_{10} > 100$, with 2.6% more errors on switch than on repeat trials. No other main effects or interactions were significant, Fs $\leq$ 4.16, ps $\geq$ .058 and Bayes factors less than 0.21.

3.4.3 Deary-Liewald task

3.4.3.1 Morningness-Eveningness Questionnaire demographics

The MEQ results for the sample (n = 123) used for the Deary-Liewald analysis were as follows: 'definite morning': n = 1 (0.8%), 'moderate morning': n = 16 (13.0%), 'intermediate': n = 74 (60.2%), 'moderate evening': n = 29 (23.6%), 'definite evening': n = 3 (2.4%). In the following analysis, I analysed MEQ as a continuous variable.

3.4.3.2 Response latency and error analysis

Latencies were analysed with a three-way ANOVA with task (SRT vs. CRT), session time (morning vs. noon vs. afternoon) and MEQ score (as a continuous variable). Results are shown in Table 3.5 and Figure 3.4.

Table 3.5

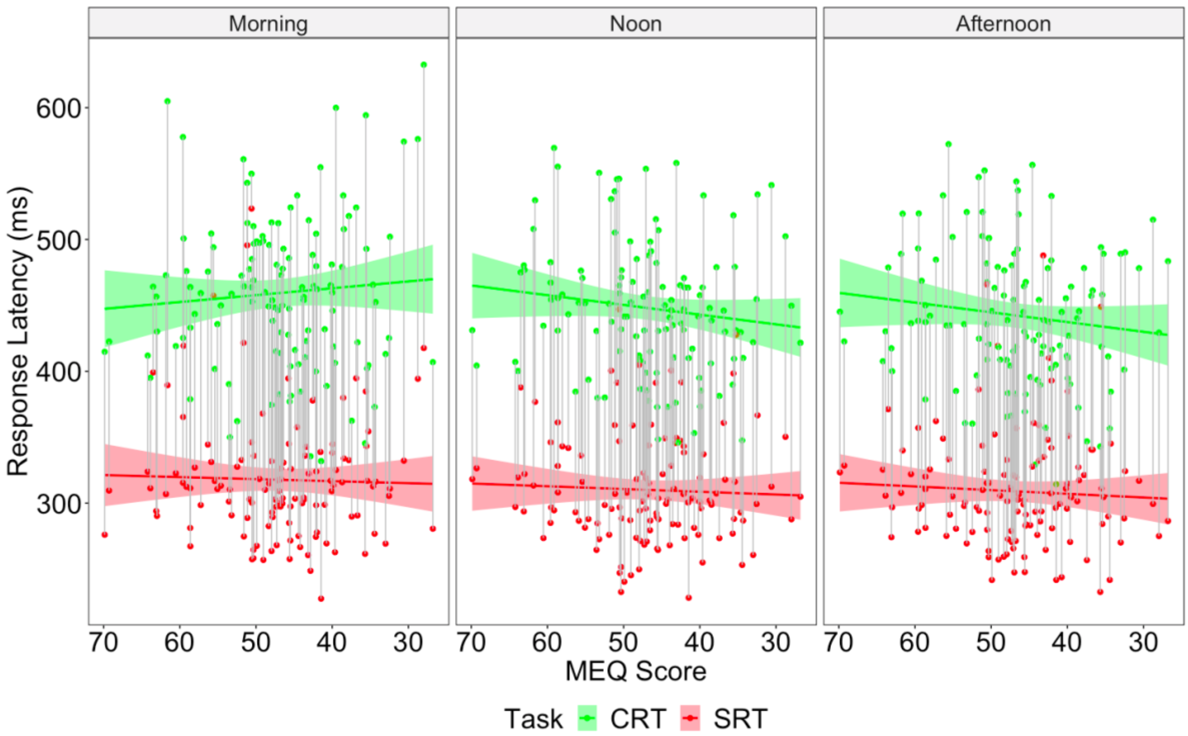
Deary-Liewald task. Analysis of Variance performed on response latencies, with task (SRT vs. CRT), session time (morning vs. noon vs. afternoon) and Morningness-Eveningness Questionnaire (MEQ) score

Effect	df_1	df_2	MSE	F	η_p^2	p	BF_{10}
Task	**1**	**121**	**2,451**	**1,427.83**	**.92**	**<.001**	**>100**
Session time	**2**	**242**	**1,231**	**8.83**	**.07**	**<.001**	**39.87**
MEQ	1	121	8,040	.55	<.01	.538	.026
Task × Session time	2	242	814	1.26	.01	.399	.059
MEQ × Task	1	121	2,451	.05	<.01	.826	.129
MEQ × Session time	2	242	1,231	2.33	.02	.174	.166
MEQ × Task × Session time	2	242	814	2.77	.02	.151	.235

Note. P values were corrected via the Benjamini-Hochburg procedure; significant effects are bolded.

Figure 3.4

Deary-Liewald task. Mean response latencies for Simple Reaction Time (SRT; red) and Choice Reaction time (CRT; green) tasks, dependent on Morningness-Eveningness Questionnaire score (MEQ; x-axis) and session time (from left to right: Morning, Noon, Afternoon session)



Note. X axis is reversed so that scores toward the left indicate a tendency toward an ‘early’ chronotype, and scores toward the right a ‘late’ chronotype. Dots represent participants and conditional means, with a grey line connecting the conditional means for a participant.

I found a highly significant effect of task, with latencies being 138ms faster in SRT trials than in CRT trials (SRT: 312ms; CRT: 450ms). Furthermore, I found a significant effect of session time, with latencies slower in the morning session (389ms) than in the noon session (379ms); $t(121) = 2.90$, $p = .012$, and the afternoon session (376ms); $t(121) = 3.89$, $p < .001$, but no significant difference between the latter two, $t(121) = 1.15$, $p = .487$. Session time affected both SRT and CRT in a very similar way (SRT: 318ms, 310ms, 309ms; CRT: 459ms, 448ms, and 442ms, for the morning, noon and afternoon sessions respectively). There was no main effect of MEQ score, nor were there any significant interactions. A parallel ANOVA conducted on error scores showed only a significant effect of task, $F(1, 121) = 189.80$, $MSE = 25$, $\eta p^2 = .611$, $p < .001$, $BF_{10} > 100$, with fewer errors on SRT trials (0.8%) than on CRT trials (5.9%.) No other main effects or interactions were significant, $F\text{s} \leq 1.47$, $p\text{s} \geq .505$ and Bayes factors less than 0.19.

3.5 Discussion

To recap, the aim of my study was to explore whether consideration of time of testing along with an individual's chronotype is required when designing studies of cognitive inhibition. For example, would I need to consider controlling synchrony effects when comparing cognitive inhibition between monolinguals and bilinguals? To achieve this, I conducted an online study with young adults which aimed to establish whether a synchrony effect can be found in tasks which measure inhibitory control, response suppression, task switching, as well as processing speed. I asked participants to repeatedly complete the Faces task and Deary-Liewald task across three times of the day (morning, noon and afternoon). A synchrony effect would be demonstrated via a three-way interaction between the inhibitory component, MEQ score and time of testing. Considering prior research (e.g., May & Hasher, 1998; Rothen & Meier, 2016, Hahn et al., 2012), one could expect that the synchrony effect emerges for measures of inhibitory control, response suppression and task switching, but that processing speed is not affected.

The Faces task measured three cognitive processes: response suppression, inhibitory control, and task switching. I found highly significant corresponding main effects (20, 27, and 20ms respectively), demonstrating that this task successfully captured these three aspects of cognitive inhibition. In addition, I found a main effect of session time, with slower latencies in the morning than in the noon and afternoon sessions (e.g., average latencies of 341, 323 and 324ms in the task switch analysis which included all trials). This pattern is expected on the observation that young adults tend towards a 'late' chronotype, and so are specifically disadvantaged in the morning session compared to the later sessions. Finally, for all three components of inhibition, I found an interaction between chronotype (MEQ) and session time. In Figures 3.1-3.3, this pattern appears as an effect of chronotype on latencies which is specific to the morning session (positive slope) but absent in the noon and afternoon sessions (horizontal slope). This pattern could be interpreted in line with the claim (see above) that young adults, due to their predominantly 'late' chronotype, are specifically affected in morning sessions, and here is where chronotype impacts latencies. By contrast, perhaps performance is at ceiling in the noon/afternoon sessions and hence, here chronotype is less relevant. However, a follow-up analysis of this pattern showed only partial statistical support for the possibility that the morning session was particularly affected by chronotype.

Of central interest was the presence or absence of a three-way interaction between a component of inhibition, chronotype, and session time, which if found, would indicate a synchrony effect. Here, the finding was clearly negative: in all three components of inhibition, this interaction was very far from significance. Visually, Figures 3.1-3.3 also suggest that in all analyses and testing sessions, the two lines representing the effect of cognitive inhibition were virtually parallel. This suggests that the efficiency of inhibitory control, response suppression and task switching remain stable across the day. Hence, my results are not in line with previous research which had reported significant effects of synchrony (e.g., Facer-Childs, Boiling & Balanos, 2018; May, 1999; May & Hasher, 1998; Ngo & Hasher, 2017). Instead, my study is consistent with studies which had also implemented a repeated-measure design and observed null findings (e.g., Barclay & Myachykov, 2017; Matchock & Mordkoff, 2009).

Furthermore, my results are inconsistent with Lustig et al.'s (2007) assertion that synchrony effects are strongest in tasks which involve heavy use of inhibition and absent in processes which rely on automatic responses. To capture a 'pure' measure of processing speed with only minimal or no cognitive inhibition, I included the Deary-Liewald task in all testing sessions. As in the Faces task, my results suggested no synchrony effects in Deary-Liewald task performance. On one hand, this result aligns with previous studies which similarly reported no synchrony effects in tasks that do not include an inhibition component (e.g., Lara, Madrid & Correa, 2014; May, 1999; Song & Stough, 2000). On the other hand, my results showed that response latencies were slowest in the morning session, a finding which mirrors the one from the Faces task. This result suggests that processing speed is susceptible to testing time, even in tasks which rely on automatic processing, and hence, contradicts research which has reported stable processing speed across the day (e.g., May & Hasher, 1998, 2017). Overall, I failed to find a clear dissociation in performance between tasks which rely on cognitive inhibition (the Faces task) and those which provide a ‘pure’ measure of processing speed (the Deary-Liewald task): both tasks exhibited time of day effects but neither showed synchrony effects. Therefore, one could conclude from my findings that researchers do not need to account for synchrony effects when designing studies involving group comparisons of inhibition and processing speed. At most, my findings imply that researchers may need to consider addressing time of day effects when comparing processing speed between groups.

An aspect of my study which may explain why I found no synchrony effects relates to the recruitment of young adults. In my study, the mean age of participants was approximately 20 years old. It is generally agreed among researchers that cognition tends to peak in young adulthood and then steadily declines across middle adulthood to late adulthood (e.g., Deary et al., 2009). For that reason, one could argue that my findings reflect young adults’ broad window of optimal performance and in turn, demonstrate their cognitive flexibility. Indeed, May and Hasher (2017) highlighted the cognitive flexibility of young ‘neutral types’. However, my study suggests this flexibility could be extended to a whole spectrum of chronotypes. Considering my findings, the next logical step would be to explore synchrony effects in inhibition for older adults. I can assume that older adults are less cognitively flexible than young adults and therefore, synchrony effects are more likely to emerge in this age group. Prior research appears to support this possibility (e.g., Borella et al., 2010; May & Hasher, 1998, 2017). Therefore, a possibility is that the alignment of an individual's chronotype and time of testing does not need to be considered when designing studies which compare inhibition performance between two groups of young adults. However, this may not be the case when comparing performance between young and older adults or comparing performance between two groups of older adults. Then again, it should be considered that the number of studies which have compared synchrony effects between young and older individuals is relatively small. To add to this, as discussed in Chapter 1, the magnitude of the difference in inhibition between the two age groups has been questioned. Nevertheless, I investigate the possibility of the synchrony effect emerging more clearly among the older population in the next chapter.

Alternatively, the absence of a synchrony effect could be explained by my sampling of participants. Unlike in studies reporting positive findings, I chose to not pre-select participants based on their chronotype and I found that more than 60% of my sample is classified as ‘neutral’ according to the conventional categorisation. The overall profile of chronotypes in my sample matches the one found in earlier studies regarding young adults (e.g., May & Hasher, 1998) but a recent study by May and Hasher (2017) found that young ‘neutral types’ are insensitive to synchrony effects. Hence, the inclusion of a large proportion of 'neutral' chronotypes might make it more difficult to detect a synchrony effect should it exist. A factor which potentially contributes to the lack of participants with ‘extreme’ chronotypes is that the majority (70%) of participants in the Faces task were females. Given that young adults generally show a tendency towards eveningness but females are less likely than males to identify as 'evening types' (e.g., Randler & Engelke, 2019), this could have led to a preponderance of ‘neutral’ chronotypes in my sample. The multi-sessional nature of my study could have additionally deterred individuals with 'extreme' chronotypes from participating in my study, or they were more likely not to complete all the critical sessions than ‘neutral’ types. For that reason, my decision to ask participants to attend multiple sessions may have resulted in a sampling bias. It is possible that the recruitment of more participants who are classified as a ‘morning’ or ‘evening' type would have revealed a potential synchrony effect more clearly.

In my study, the ‘late’ (afternoon) sessions occurred between 4pm and 5pm (GMT time). It is possible that synchrony effects would have been found with substantially later testing times; for instance, participants categorised as a ‘definitely evening’ type may not have been able to provide their optimal performance between 4pm and 5pm (GMT time), but perhaps only later in the evening. Studies on time of day and synchrony effects have implemented large variations in testing times, with afternoon sessions starting after 12pm (Murphy et al., 2007), taking place between 3pm - 6:30pm (Intons-Peterson et al., 1998), 4pm - 5pm (Yang et al., 2007; Yoon et al., 1999), or between 4:15pm - 5:15pm (Hasher et al., 2002), or at 5pm (Bugg et al., 2006; West et al., 2002). I chose to implement a similar timing of sessions to studies which have previously reported significant synchrony effects (e.g., May & Hasher, 1998; Hasher et al., 2002; Yang et al., 2007) and adding to this, my testing times align with a typical working day (i.e., start work between 8am and 9am, lunchtime between 12pm and 1pm and finish work between 4pm and 5pm). It is also worth noting that there are prior studies which implemented afternoon/evening sessions later than mine, but which nonetheless failed to detect synchrony effects. For instance, Matchock and Mordkoff (2009) asked participants to complete the ANT task at four time points (8am, 12pm, 4pm and 8pm) and showed that task performance for 'evening types' did not differ between the 8am condition and 8pm condition. To add to this, considering that in my study morning and afternoon sessions were at least seven hours apart, I should still expect 'morning types' to perform noticeably better in the morning session than in the afternoon session, and vice versa for 'evening types'. Hence, I consider it unlikely that the lack of synchrony effect in my findings can be explained by my choice of session timings. Nevertheless, I acknowledge that future studies should aim to investigate synchrony effects with later testing times.

In my study, experimental sessions were scheduled based on external time, and hence, I did not account for individual differences in sleep-wake habits. As a consequence, for some participants, the sessions in my study may have conflicted with their sleep-wake habit and thus, their performance might have been disrupted. Schmidt et al. (2012) explored whether the magnitude of the synchrony effect depended on whether or not sessions were scheduled based on participants' sleep-wake habits. In their study, older 'morning types' and young 'evening types' completed the Stroop and Psychomotor vigilance tasks, under both external and internal timing conditions (for the latter, the timings of the sessions depended on the participants' sleep-wake time). Significant synchrony effects were reported for the external time condition but became non-significant for the internal time condition. These findings suggest that an individual's sleep-wake habits can be a confounding factor in previous synchrony effect studies. Regarding my own results, the significant time of day effects could be explained by my use of external timings: the morning session may have encouraged participants to wake up earlier than they usually do and as a result, the timing of this session may have elicited slow task performances. Future research could replicate my study but instead adapt the session timings to participants' individual sleep-wake schedules.

A related limitation of my study is associated with my sleeping measures. Unlike in most previous studies, I attempted to ensure that lack of sleep did not impact cognitive task performance. Here, I assumed that sleeping for at least five hours the night before a critical session is sufficient for an individual. This measure of sleepiness could be seen as flawed due to individual differences in perceived sufficient hours of sleep (i.e., the optimal number of hours of sleep varies from person to person; Chaput, Dutil & Sampasa-Kanyinga, 2018). For example, some people may be able to be fully alert with five hours of sleep whereas others require eight hours of sleep. As a consequence, it may be the case that some of my participants have reported sleeping for more than 5 hours but still felt like they had a lack of sleep. Adding to this, I did not ask questions regarding their quality of sleep. Undergoing enough hours of sleep does not necessarily translate into good quality sleep. For example, some participants may have slept for 8 hours but they had disrupted sleep and in turn, did not feel refreshed after waking up. Indeed, poor quality of sleep has been associated with poor cognitive performance (e.g., Wilckens et al., 2014; della Monica et al., 2018). Considering these limitations, arguably, I cannot entirely rule out the sleep-related influences in my study. Future studies may consider adding questions which assess a participant's average and reported number of hours of sleep and their quality of sleep.

3.6 Conclusion

To summarise, my results suggest that the interplay between chronotype and time of testing did not impact young adults' response suppression, inhibitory control, or task switching, nor their processing speed. Failure to report a synchrony effect in measures of processing speed is in line with prior studies which likewise reported no synchrony effects in tasks which require automatic response. However, the lack of synchrony effects in activities which heavily rely on cognitive inhibition contradicts previous research which had suggested effects of this type. For that reason, my findings indicate that researchers may not need to be concerned about these effects when comparing cognition inhibition and processing speed between groups. At most, the significant main effect of session time suggests that it may be best to test young adults from noon onwards (which I do in Chapter 6). Then again, I speculate that the synchrony effect could be genuine but my inability to capture it could be attributed to my recruitment of young adults and sampling of participants which entailed the inclusion of a large proportion of 'neutral' chronotypes. I attempt to explore the former possibility in the next chapter where I essentially replicate the design of this study but with older adults.

Chapter 4: Synchrony effects among older adults

4.1 Chapter aims

To concisely summarise the previous chapter, I found no significant synchrony effects in two forms of inhibition (i.e., response suppression and inhibitory control), task switching and processing speed among young adults. One potential explanation introduced in the discussion section is the fact that young adults may be resilient towards synchrony effects. Therefore, in this chapter, I investigated whether this effect can be found more clearly among older adults, using mostly the same methodology as I did in Chapter 3. The rationale of why there may be age-related differences in results is that young adults are at a cognitive peak whereas older adults are experiencing a cognitive decline. For that reason, one could argue that older adults may be more vulnerable to synchrony effects than younger adults.

4.2 Introduction

In comparison to the literature with young adults (see Chapter 3), there are noticeably fewer studies which explore synchrony effects in cognitive inhibition among older adults. As you can see in Table 4.1, the few published studies which explore synchrony effects in cognitive inhibition among this age group are between-participants designed and involved 'morning types'. This is perhaps unsurprising given that older 'evening types' are difficult to recruit since there appear to be so few of them. Therefore, by recruiting only 'morning types', the question that is explored in many of the studies included in Table 4.1 is whether older adults perform the tasks better in the morning than in the afternoon/evening (i.e., a synchrony effect). Adding to this, many of the studies included in Table 4.1 involved comparisons between the young and older adult populations. Numerous studies report significant findings in support of a synchrony effect among the older adult population but not for the young adult population (e.g., May & Hasher, 1998). Therefore, this is consistent with my speculation that older adults are more vulnerable to synchrony effects than young ones.

Then again, like the literature surrounding synchrony effects among the young adult population, findings are mixed. On the one hand, some studies including May and Hasher (1998) and Intons-Peterson et al. (1999) observed superior task performance in the morning compared to the afternoon group. On the other hand, null findings involving ANT and flanker tasks were also reported which are not in line with Lustig et al.'s (2007) assertion that synchrony effects are strongest for inhibition tasks. Considering these findings, no clear conclusions can be made on whether older 'morning types' perform tasks better in the morning than afternoon. Again, mirroring the young adult literature, research exploring synchrony effects among older adults are compromised with low sample sizes and a largely exclusive focus on participants with a non-neutral chronotype.

To my knowledge, May and Hasher's (2017) paper is the only study which explicitly explored synchrony effects in older 'neutral types'. In this study, participants undertook a battery of cognitive tasks including sentence completion, Stroop and letter comparison. These 'neutral types' attended the test session at one of three time points (i.e., 8am, 12pm or 5pm). For older adults, Stroop effects were reported to be largest in the afternoon session and hence, this result indicates a synchrony effect. However, the results suggest that the magnitude of the Stroop effect did not differ between the morning and noon groups. The authors concluded that the equivalent results demonstrate the participants' transition from being a 'neutral type' to a 'morning type'. This reasoning would suggest that individuals transitioning from one chronotype to another would have a broader optimal performance time window than those not transitioning. Nevertheless, the findings of May and Hasher's (2017) study are consistent with the narrative that older adults generally perform better in the morning than afternoon. Adding to this, the fact that the synchrony effect does not emerge among young 'neutral types' further supports the idea that the synchrony effect appears more strongly among older than young adults.

Table 4.1
Summary of previous studies on synchrony effects in inhibition among older individuals, by publication year

Reference	Design (between/ within Ss)	Sample (older participants)	Age range	Chronotype	Time window	Task	Non-behavioural measures included?	Behavioural results	Synchrony effect?
May et al. (1993), Exp. 2	between	N=22 (11 randomly assigned to AM and PM group)	66-78	characterised as 'strong morningness', probably implying screening for 'definite morning type' (MEQ scores ≥ 70)	AM: 8-9AM PM: 4-5PM	Verbatim sentence recognition from passages of text	no	Significantly poorer recognition performance in PM than AM (mainly due to a rise in 'false alarms')	✓
May & Hasher (1998), Exp. 1	between	N=48 (24 randomly assigned to AM and PM group)	62-75	screened to be of 'morning type' (MEQ score ≥ 59)	AM: 8AM PM: 5PM	Sentence completion task, requiring suppression of no-longer-relevant information	no	Priming for disconfirmed items numerically larger in PM than AM, but comparison only marginally significant	?

Reference	Design (between/ within Ss)	Sample (older participants)	Age range	Chronotype	Time window	Task	Non-behavioural measures included?	Behavioural results	Synchrony effect?
May & Hasher (1998), Exp. 2	between	N=36 (18 randomly assigned to AM and PM group)	62-75	screened to be of 'morning type' (MEQ score ≥ 59)	AM: 8AM PM: 5PM	Stop-signal; Stroop colour naming (non-computerised)	no	Stop-signal task: stopping probability and stopping times significantly lower in PM than AM Stroop: larger Stroop effect in PM than AM	✓
Li et al. (1998), Exp. 1	between	N=32 (16 randomly assigned to AM and PM group)	63-75	screened to be of 'morning type' (MEQ score ≥ 59)	AM: 8AM PM: 5PM	Reading aloud task with distracting words; recognition memory	no	No effect of time of day on reading performance	✕
Li et al. (1998), Exp. 2	between	N=32 (16 randomly assigned to AM and PM group)	63-75	screened to be of 'morning type' (MEQ score ≥ 59)	AM: 8AM PM: 5PM	Reading aloud task with distracting words; recall memory	no	No effect of time of day on reading performance	✕

Reference	Design (between/ within Ss)	Sample (older participants)	Age range	Chronotype	Time window	Task	Non-behavioural measures included?	Behavioural results	Synchrony effect?
Intons-Peterson et al. (1999), Exp. 1	between	N=42 (20 assigned to 'optimal' group, 22 assigned to 'nonoptimal' group)	60-90	5 adults characterised as 'evening type', the rest were 'morning type'	AM: before 10:30AM PM: at or after 3PM	Verbal false memory paradigm	no	Non-optimal group recalled/and recognised more non-studied lures (indicating less efficient inhibition of nonrelevant items) than the optimal group	✓
West et al. (2002)	within	N=20 (10 randomly assigned to AM-PM or PM-AM testing order)	mean age = 72.8	predominantly 'neutral' or 'moderate morning'; not screened based on MEQ score	AM: 9AM PM: 5PM	Four-box task (spatial working memory task with distractors and n-back manipulation)	no	Numerically larger distractor effects in PM than AM, but performance difference not directly assessed statistically	?

Reference	**Design (between/ within Ss)**	**Sample (older participants)**	**Age range**	**Chronotype**	**Time window**	**Task**	**Non-behavioural measures included?**	**Behavioural results**	**Synchrony effect?**
Rowe, Hasher & Turcotte (2009)	between	N = 56 (28 randomly assigned to AM and PM group)	60 - 76	screened to be of 'morning type' (MEQ score ≥ 59)	AM: 8AM or 9AM PM: 4PM or 5PM	Corsi Task (measure of visual spatial working memory) with participants completing a high (short list first) or low interference format (large list first)	no	Better performance for AM than PM group, for the low-interference format only	?

Reference	Design (between/ within Ss)	Sample (older participants)	Age range	Chronotype	Time window	Task	Non-behavioural measures included?	Behavioural results	Synchrony effect?
Borella et al. (2010)	between	N=40 (20 randomly assigned to AM and PM group)	60-78	not measured/reported	AM: 8-11AM PM: 2-5PM	Stroop colour naming (single trial)	no	Stroop interference (incongruent-control) significantly larger for PM than AM group. Stroop facilitation (congruent-control): no difference between AM/PM	✓
Knight & Mather (2013)	between	N=32 (16 randomly assigned to AM and PM group)	65-85	not measured/reported	AM: 8-10AM PM: 2-5PM	Attentional Network Task (ANT)	no	Significant difference in 'alerting' between the AM and PM group; no difference in 'orienting' or 'executive attention'	?

Reference	Design (between/ within Ss)	Sample (older participants)	Age range	Chronotype	Time window	Task	Non-behavioural measures included?	Behavioural results	Synchrony effect?
Anderson et al. (2014)	between	N=32 (16 randomly assigned to AM and PM group)	60-87	predominantly 'moderate morning'; not screened based on MEQ	AM: 8:30-10:30AM PM: 1-5PM	Arrow-flanker	fMRI	Descriptively but not statistically significantly larger Flanker RT effect in PM than in AM older adults	?
May & Hasher (2017)	between	N=72 (24 randomly to assigned to AM, noon, and PM group)	60-74	predominantly 'neutral' (mean MEQ score of 52)	AM: 8:00 Noon: 12:00 PM: 5:00	Sentence completion; Stroop	no	Sentence completion: significant distractor priming in AM and PM but not at midday. Stroop: largest Stroop effect in PM. However, no significant difference in effect size between AM and midday.	?

Reference	**Design (between/ within Ss)**	**Sample (older participants)**	**Age range**	**Chronotype**	**Time window**	**Task**	**Non-behavioural measures included?**	**Behavioural results**	**Synchrony effect?**
Rabi et al. (2022a, 2022b)	between	N=52 (26 randomly assigned to AM and PM group)	64-88	screened to be of 'morning type' (MEQ score ≥ 59)	AM: 8-10:30AM PM: 2-5PM	Go/no-go; Flanker	EEG	Performance not affected by time of day for either task	✕

4.2.1 The present study

To recap, the main aim of my study was to clarify whether I can observe a synchrony effect in the older adult population. In the previous chapter, I was unable to observe synchrony effects among young adults. Considering that older adults are thought to possess less efficient inhibition than young adults (e.g., see the inhibitory deficit theory), I speculate that perhaps synchrony effects would emerge more clearly among this age group. However, as mentioned in Chapter 1, recent papers suggest that the age-related differences between older and young adults may have been 'greatly exaggerated' (e.g., Rey-Mermet et al., 2018; Verhaeghen, 2011). Therefore, if the conclusions from these papers are accurate, then my prediction that older adults are more susceptible to synchrony effects than young adults may be incorrect.

Like the study described in Chapter 3, my study deviates from previous studies with my use of a large sample size, the decision to treat chronotype as a continuous variable and to implement a within-participants design. The former was achieved by conducting my study online and as a result, I was able to recruit over 100 participants. The latter was achieved by asking participants to complete an MEQ and determine whether the MEQ score predicts task performance in the morning and afternoon sessions. In a secondary analysis, I restricted my sample to those individuals with ‘morning’ chronotypes (n = 61) and I directly tested whether cognitive performance was poorer in the afternoon than in the morning session.

4.3 Method

4.3.1 Participants

I recruited 159 older adults (Male = 56, Female = 103, mean age = 71.2 years old, age range = 65-85) through the Join Dementia Research (JDR) platform. JDR is a volunteer pool which consists of people who are interested in participating in dementia-related research. All participants were reported as: being monolingual, aged between 65 to 85, currently residing in the UK at the time of testing, not colour-blind, not shift-workers, having not recently travelled abroad, having not experienced sleeping problems and not having severe memory problems. Finally, all participants reported not experiencing symptoms of long-COVID. I implemented this criteria because long-COVID has been observed to be associated with brain fog which may negatively influence cognitive performance (e.g., Hugon et al., 2022). This study has received ethical approval from the University of Bristol Faculty of Life Sciences

Research Ethics Committee (code: 12121997085) and informed consent was provided by all participants.

4.3.2 Materials

The same questionnaires and tasks were used as in Chapter 3 except in the background questionnaire where I added an additional question asking whether participants think they are experiencing any symptoms of brain fog caused by COVID-19.

4.3.3 Procedure

My study consisted of three consecutive online sessions. I chose to design the study as consisting of three rather than four online sessions (like in the young adult version) and session time was not counterbalanced. This was because I was unable to send participants reminders of upcoming testing sessions. As a consequence, counterbalancing session time order and implementing four sessions may have led to high attrition rates. Like in the study described in Chapter 3, participants first underwent a practice session the day before the experiment sessions. This session consisted of an MEQ and background questionnaire and the practice versions of the two critical tasks. The time window for the morning session was between 8am and 9am (GMT time) and the time window for the afternoon session was between 4pm and 5pm (GMT time). These sessions lasted approximately 30 minutes and participants would fill out a session questionnaire and complete the two tasks. In the afternoon session, participants were debriefed after completing the Deary-Liewald task.

4.4 Results

4.4.1 Pre-processing

Participants were excluded from data analysis for the following reasons: slept for less than 5 hours before the morning session ($n = 2$), consumed too much alcohol before the critical sessions ($n = 13$) or produced an error rate higher than 25%. Here, I should note that participants were excluded from the analysis of the task in which they produced the high error rates in. For example, if a participant elicited an error rate larger than 25% in the Faces task but not on the Deary-Liewald task, they would be excluded from the analysis of the former task only. As a result, 13 participants were excluded from the Faces task analysis and

25 participants were excluded from the Deary-Liewald task analysis. For the Faces task analysis, I excluded 1 participant due to missing data and 1 participant for producing long response latencies (their average response latency was larger than two standard deviations above the mean). For the Deary-Liewald task, I excluded 2 participants due to missing data. Overall, I included 129 participants (Female = 84, Male = 45, Mean age = 70.9 years old) in the final Faces task analysis and 117 participants (Female = 76, Male = 41, Mean age = 70.9 years old) in the Deary-Liewald task analysis.

4.4.2 Faces task

4.4.2.1 Morningness-Eveningness Questionnaire

I would like to first note that I analysed MEQ as a continuous variable. However, if MEQ was treated as a categorical variable, the demographic profile would be as follows: 'definite evening': n = 0 (0.0%), 'moderate evening': n = 4 (3.1%), 'intermediate': n = 64 (49.6%), 'moderate morning': n = 55 (42.6%) and 'definite morning': n = 6 (4.7%).

4.4.2.2 Response suppression

Recall, response suppression was determined by the difference in performance between the ipsilateral and contralateral conditions for neutral trials only. I performed an ANOVA with within-participants factors, response type (ipsilateral vs. contralateral) and session time (morning vs. afternoon) and between-participant factor, MEQ score. As can be seen in Table 4.2, I found a significant main effect of response type, $F(1,127) = 29.95$, $p < .001$, with participants being 24ms faster at making an ipsilateral (524ms) than contralateral response (548ms). This finding suggests that the Faces task was successful in capturing response suppression. Additionally, I found a significant main effect of session time. Unexpectedly, I found that participants were 35ms faster in the afternoon session (518ms) than in the morning session (553ms).

Fundamentally, the ANOVA revealed no significant three-way interaction between MEQ, session time and response type, $F(1, 127) = 0.49$, $p = .724$, and thus, indicating failure to report a significant synchrony effect in response suppression. Additionally, I used Bayesian statistics to investigate how strong the evidence is in support of this null finding. I observed a

Bayes factor of 0.21 which implies 'moderate' evidence in favour of a non-significant three-way interaction.

Table 4.2

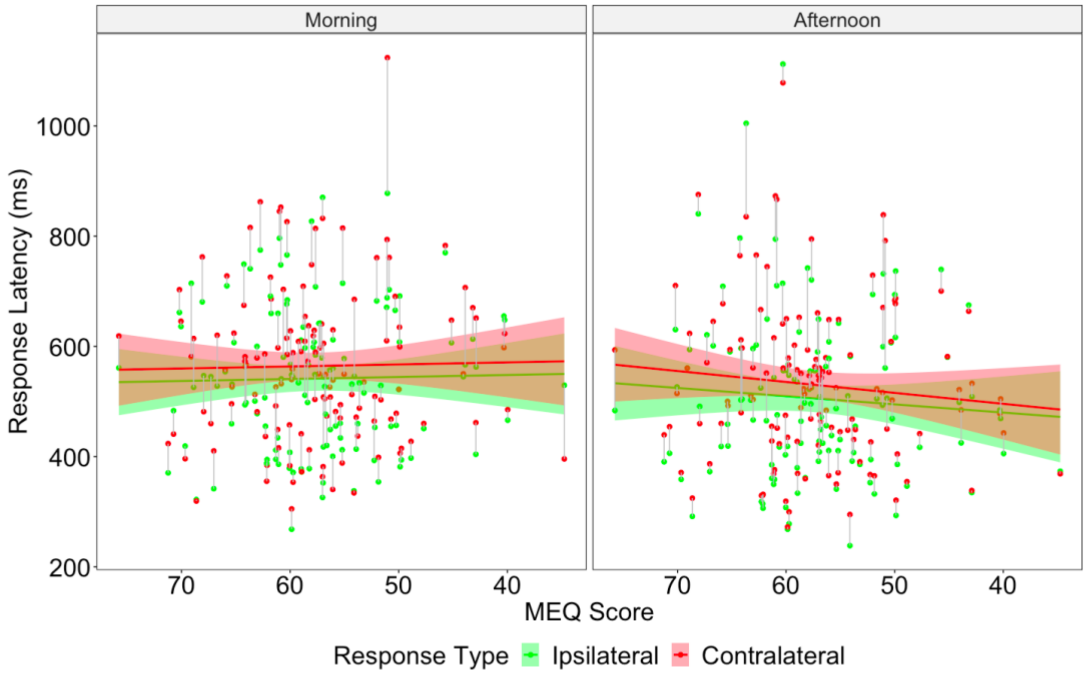
Faces task. Analysis of Variance performed on response latencies, with response type (contralateral vs. ipsilateral), session time (morning vs. afternoon) and Morningness-Eveningness Questionnaire (MEQ) score.

Effect	df_1	df_2	MSE	F	η_p^2	p	BF_{10}
Response type	**1**	**127**	**2,410**	**29.95**	**.191**	**<.001**	**>100**
Session time	**1**	**127**	**5,477**	**28.41**	**.183**	**<.001**	**>100**
MEQ	1	127	67,792	0.19	.001	.724	0.40
Response type × Session time	1	127	886	0.13	<.01	.724	0.14
MEQ × Response type	1	127	2,410	0.20	<.01	.724	0.17
MEQ × Session time	1	127	5,477	5.36	.04	.052	17.79
MEQ × Response type × Session time	1	127	886	0.49	<.01	.724	0.21

Note. P values were corrected via the Benjamini-Hochburg procedure; significant effects are bolded.

Figure 4.1

Faces task. Mean response latencies for ipsilateral (green) and contralateral (red) trials, dependent on Morningness-Eveningness Questionnaire score (MEQ; x-axis) and session time (from left to right: Morning and Afternoon session)



Note. X axis is reversed so that scores toward the left indicate a tendency toward an 'early' chronotype, and scores toward the right a 'late' chronotype. Dots represent participants and conditional means, with a grey line connecting the conditional means for a participant.

Parallel analysis on error rates indicates no significant main effects nor any significant interactions (all *F*s < 6.97, *p*s > .065 and Bayes factors less than 3.95).

4.4.2.3 Inhibitory control

Inhibitory control was calculated by the difference in performance between the congruent and incongruent conditions for the ipsilateral trials only. Again, I performed a mixed factors ANOVA but in this case with factors, congruency (congruent vs. incongruent), session time (morning vs. afternoon) and MEQ score. Table 4.3 displays the results of this ANOVA. As can be seen, I observed a significant main effect of congruency with response latencies being 19ms faster for congruent (503ms) than incongruent (522ms) trials. Like the results for the response suppression analysis, I observed a significant main effect of session time with faster responses in the afternoon (527ms) than morning session (498ms).

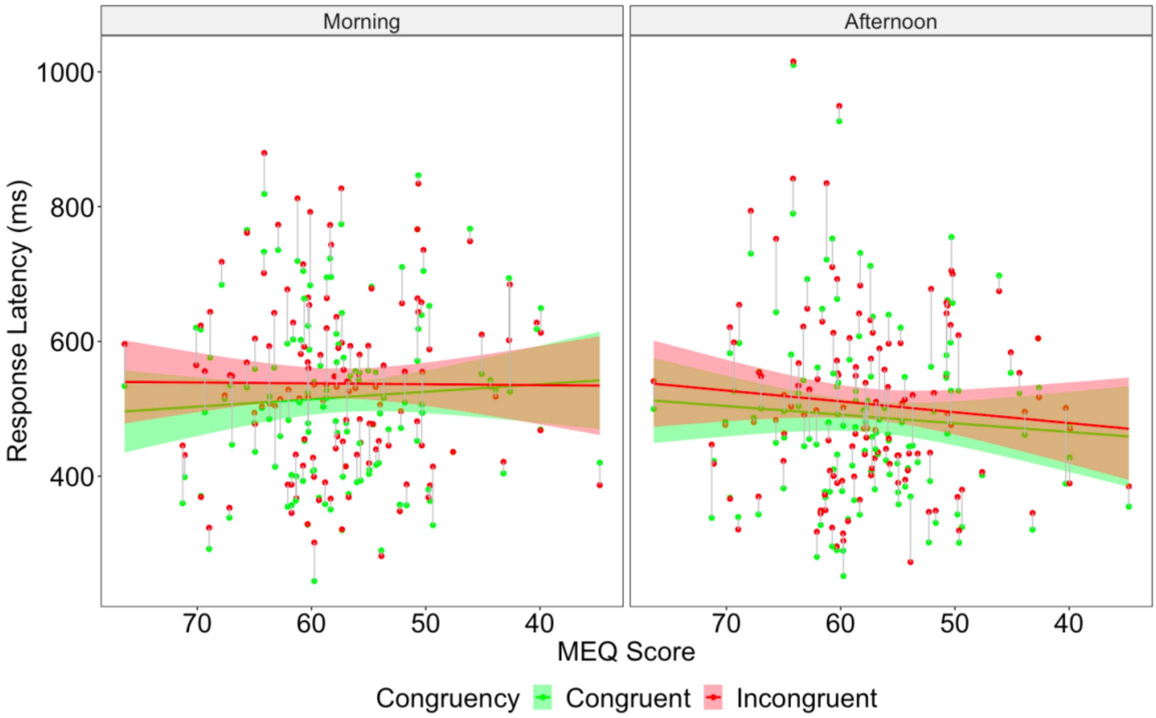
Table 4.3
Faces task. Analysis of Variance performed on response latencies, with congruency (congruent vs. incongruent), session time (morning vs. afternoon) and Morningness-Eveningness Questionnaire score (MEQ)

Effect	df_1	df_2	MSE	F	η_p^2	p	BF_{10}
Congruency	**1**	**127**	**983**	**50.93**	**.286**	**<.001**	**>100**
Session time	**1**	**127**	**4,181**	**27.06**	**.176**	**<.001**	**>100**
MEQ	1	127	61,013	0.09	<.01	.759	0.44
Congruency x Session time	1	127	698	0.11	<.01	.759	0.13
MEQ x Congruency	1	127	983	4.22	.032	.074	0.51
MEQ x Session time	**1**	**127**	**4,181**	**5.77**	**.043**	**.041**	**52.03**
MEQ x Congruency x Session time	1	127	698	1.86	.014	.246	0.27

Note. P values were corrected via the Benjamini-Hochburg procedure; significant effects are bolded.

Figure 4.2

Faces task. Mean response latencies for congruent (green) and incongruent (red) trials, dependent on Morningness-Eveningness Questionnaire score (MEQ; x-axis) and session time (from left to right: Morning and Afternoon session)



Note. X axis is reversed so that scores toward the left indicate a tendency toward an 'early' chronotype, and scores toward the right a 'late' chronotype. Dots represent participants and conditional means, with a grey line connecting the conditional means for a participant.

As you can see in Table 4.3, I found a significant interaction between MEQ and session time and Figure 4.2 indicates that MEQ may influence response latencies in the afternoon session only. However, with further investigation, I failed to uncover a significant main effect of MEQ in the morning and afternoon sessions (*F*s < .79 and *p*s > .375). Critically, I was unable to report a significant three-way interaction between congruency, session time and MEQ. Bayesian statistics reveal that there is 'moderate' support in favour of this null finding (BF_{10} = 0.27)

The three-way ANOVA for error rates revealed no significant main effects or interactions (all *F*s < 3.81, *p*s > .233 and Bayes factor less than 0.63).

4.4.2.4 Task switching

A mixed ANOVA with within-participants factors, task switch (task switch vs. task repeat) and session time (morning vs. afternoon) and between-participants factor MEQ score was used to analyse the task switching data. The output of this ANOVA can be seen in Table 4.4. I found a significant main effect of task switch with a 35ms difference between task repeat (511ms) and task switch (546ms) trials. Similarly, I observed a significant session time effect with participants performing, on average, faster in the afternoon (512ms) than morning (545ms).

In terms of interactions, as can be seen in Table 4.4, the ANOVA uncovered a significant interaction between MEQ and session time. Similar to the inhibitory control analysis, post-hoc analysis suggests that MEQ did not significantly impact response latencies in the morning and afternoon sessions (*F*s < 1.23 and *p* > .270). Furthermore, the Bayesian statistics indicate 'moderate' evidence in favour of a null finding (BF_{10} = 0.25).

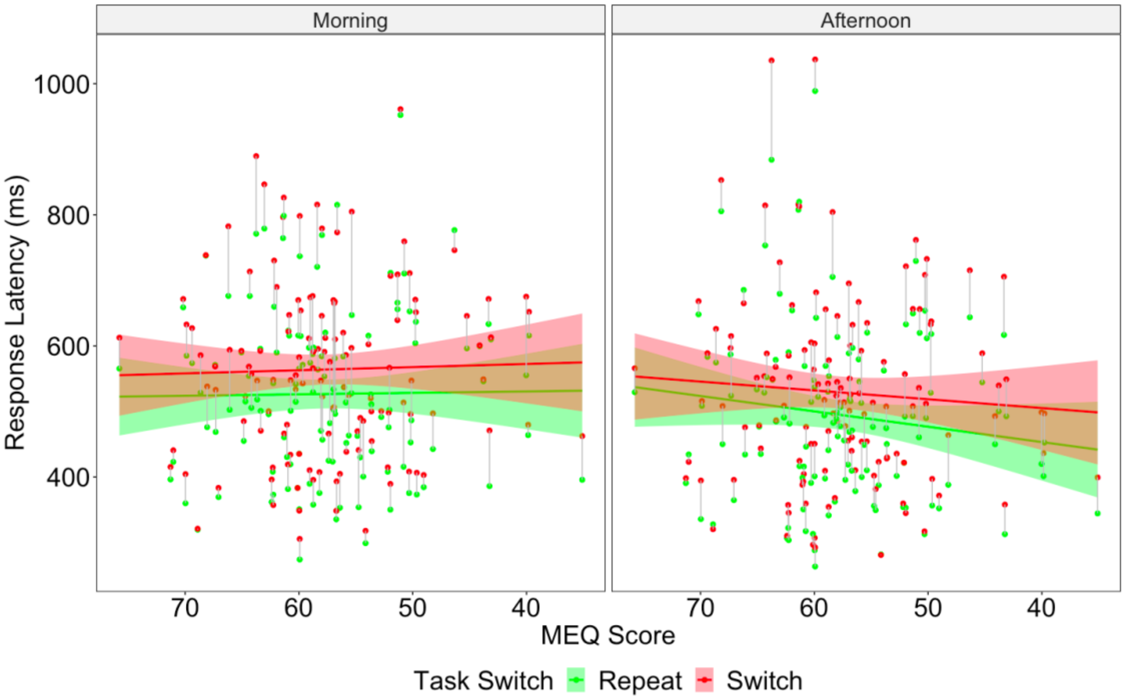
Table 4.4
Faces task. Analysis of Variance performed on response latencies, with task switch (task switch vs. task repeat), session time (morning vs. afternoon) and Morningness-Eveningness Questionnaire score (MEQ)

Effect	df_1	df_2	MSE	F	η_p^2	p	BF_{10}
Task Switch	**1**	**127**	**834**	**198.81**	**.610**	**<.001**	**>100**
Session time	**1**	**127**	**4,250**	**34.36**	**.213**	**<.001**	**>100**
MEQ	1	127	64,031	0.21	<.01	.650	0.47
Task Switch x Session time	1	127	406	0.07	<.01	.463	0.14
MEQ x Task Switch	1	127	834	3.28	.025	.127	0.33
MEQ x Session time	**1**	**127**	**4,250**	**7.71**	**.057**	**.015**	**>100**
MEQ x Task Switch x Session time	1	127	406	2.26	.017	.189	0.25

Note. P values were corrected via the Benjamini-Hochburg procedure; significant effects are bolded.

Figure 4.3

Faces task. Mean Response Latencies for task repeat (green) and task switch (red) Trials, dependent on Morningness-Eveningness Questionnaire score (MEQ; x-axis) and session time (from left to right: Morning and Afternoon session)



Note. X axis is reversed so that scores toward the left indicate a tendency toward an 'early' chronotype, and scores toward the right a 'late' chronotype. Dots represent participants and conditional means, with a grey line connecting the conditional means for a participant.

The ANOVA on errors only revealed a significant main effect of task switch with average error rates for task switch trials (4.90%) being higher than error rates for task repeat trials (3.03%), $F(1, 127) = 94.82$, $MSE = 5$, $\eta p^2 = .427$, $p < .001$, $BF_{10} > 100$.

4.4.2.5 Analysis restricted to 'morning' types only

As outlined above, my sample constitutes a random subset of the population of older adults. As shown in the section 'Morningness-Eveningness Questionnaire demographics', there were virtually no 'evening' types but about half of the participants were in the 'intermediate' category. It could be argued that the lack of an effect corresponding to synchrony in my findings is perhaps not surprising, given that half of the participants were of the 'neutral type'. To test this argument, I conducted an additional analysis with response latencies in which I restricted participants to 'morning' types (MEQ scores ≥ 59, n = 61). This allowed me to omit the MEQ variable from the analysis and hence, to simplify my design. The three components of cognition were analysed via separate ANOVAs, with component (treatment vs. baseline) and session time as within-participants factors. For response suppression, this analysis showed an effect of response type, $F(1, 60) = 18.94$, $p < .001$, $\eta_p^2 = .240$, $BF_{10} = 67.18$, and of session time, $F(1, 60) = 5.70$, $p = .030$, $\eta_p^2 = .087$, $BF_{10} = 25.08$, but critically no interaction, $F(1, 60) = 0.90$, $p = .345$, $\eta_p^2 = .015$, $BF_{10} = 0.24$. For inhibitory control, a main effect of congruency was found, $F(1, 60) = 44.02$, $p < .001$, $\eta_p^2 = .423$, $BF_{10} > 100$, and of session time, $F(1, 60) = 5.68$, $p = .030$, $\eta_p^2 = .086$, $BF_{10} = 46.29$, but there was no interaction, $F(1, 60) = 1.01$, $p = .318$, $\eta_p^2 = .017$, $BF_{10} = 0.24$. Finally, for switching, there was a main effect of switch, $F(1, 60) = 75.24$, $p < .001$, $\eta_p^2 = .556$, $BF_{10} > 100$, and of session time, $F(1, 60) = 6.51$, $p = .020$, $\eta_p^2 = .098$, $BF_{10} > 100$, but, again, there was no interaction, $F(1, 60) = 0.62$, $p = .436$, $\eta_p^2 = .010$, $BF_{10} = 0.19$. In summary, even in analyses restricted to the ‘morning’ type individuals, session time never interacted with chronotype, with Bayesian statistics providing ‘moderate’ support for the null finding. Hence, no evidence was found that for ‘morning’ types, inhibitory performance was poorer in the afternoon than in the morning session.

4.4.3 Deary-Liewald task

4.4.3.1 Morningness-Eveningness Questionnaire demographics

I would like to first note that I analysed MEQ as a continuous variable. However, if MEQ was treated as a categorical variable, the demographic profile would be as follows: 'definite evening': n = 0 (0%), 'moderate evening': n = 3 (2.3%), 'intermediate': n = 60 (51.2%), 'moderate morning': n = 48 (41.0%) and 'definite morning': n = 6 (5.1%).

4.4.3.2 Final analyses

Like the analysis for the Faces task, I used an ANOVA to analyse the data for the Deary-Liewald task with factors, task (SRT vs. CRT), session time (morning vs. afternoon) and MEQ score. As can be seen in Table 4.5, I found significant main effects of task and session time. The former effect suggests that participants responded 259ms faster on SRT trials (373ms) than on CRT ones (632ms). The latter effect implies that average response latencies were slower in the morning (507ms) than afternoon session (498ms). I found no other significant effects or interactions (all *p*s > .317).

The error rate analysis, involving a three-way ANOVA, revealed no significant main effects or interactions (all *F*s < 5.58, *p*s > .139 and Bayes factors less than 1.56).

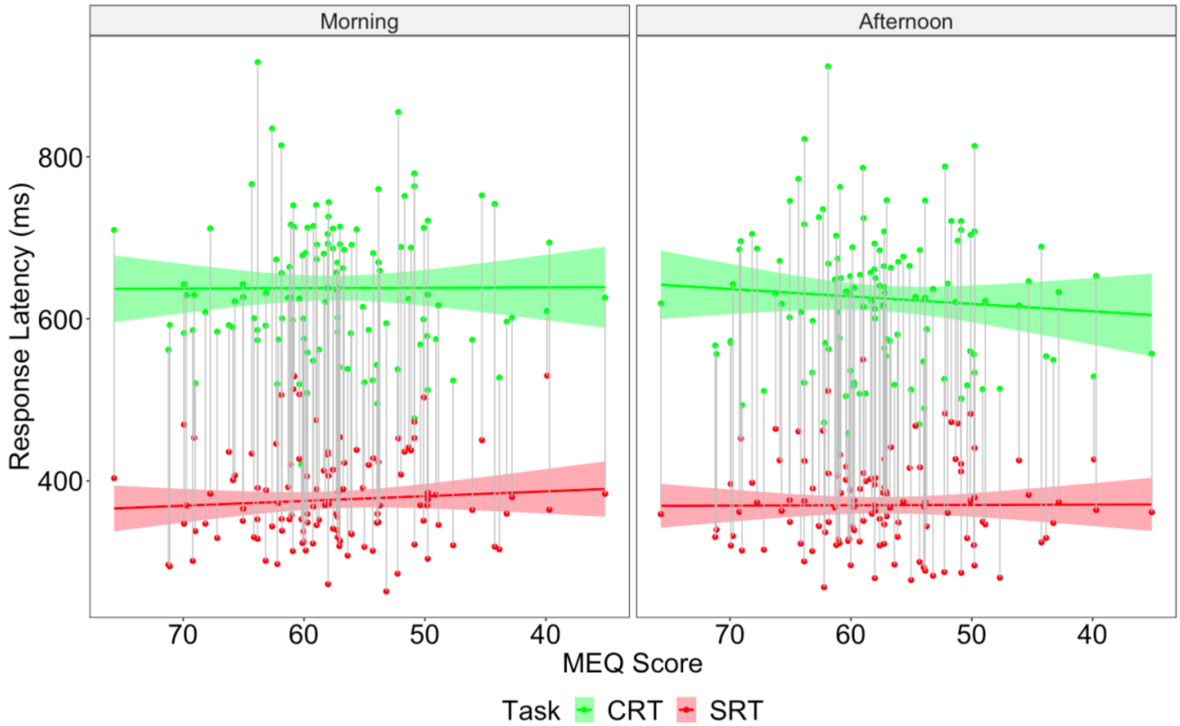
Table 4.5
Deary-Liewald task. Analysis of Variance performed on response latencies, with task (SRT vs. CRT), session time (morning vs. afternoon) and Morningness-Eveningness Questionnaire score (MEQ)

Effect	df_1	df_2	MSE	F	η_p^2	p	BF_{10}
Task	**1**	**115**	**4,287**	**1,821.45**	**.94**	**<.001**	**>100**
Session time	**1**	**115**	**1,469**	**7.33**	**.06**	**.027**	**1.11**
MEQ	1	115	14,148	0.01	<.01	.903	0.25
Task × Session time	1	115	665	1.48	.01	.395	0.17
MEQ × Task	1	115	4,287	0.90	.01	.483	0.39
MEQ × Session time	1	115	1,469	2.26	.02	.317	0.34
MEQ × Task × Session time	1	115	665	0.40	<.01	.614	0.24

Note. P values were corrected via the Benjamini-Hochburg procedure; significant effects are bolded.

Figure 4.4

Deary-Liewald task. Mean response latencies for Simple Reaction Time (SRT; red) and Choice Reaction time (CRT; green) tasks, dependent on Morningness-Eveningness Questionnaire score (MEQ; x-axis) and session time (from left to right: Morning and Afternoon session)



Note. X axis is reversed so that scores toward the left indicate a tendency toward an 'early' chronotype, and scores toward the right a 'late' chronotype. Dots represent participants and conditional means, with a grey line connecting the conditional means for a participant.

4.5 Discussion

In this chapter, I compared older adults' morning performance in the Faces and Deary-Liewald tasks with their afternoon performance. I used a similar design in this study to the one used the in previous study. Like the young adults version of the study, I treated MEQ as a continuous variable and performed the study online. Assuming that older adults possess worse inhibition than young adults, as advocated by the inhibitory deficit theory, I expected the synchrony effect in the indices of the Faces task (i.e., inhibitory control, response suppression and task switching) to arise clearly with this age group. However, I predicted this effect to not emerge for performance in the Deary-Liewald task.

To summarise, the results of this study mirror the results observed in Chapter 3 in numerous ways. Like the results reported in the previous chapter, the Faces task successfully captured the two components of inhibition as well as task switching, with significant effects. To add to this, I found that these cognitive components did not significantly interact with session time and chronotype. This suggests that I was unable to capture synchrony effects in inhibition among older adults. This pattern of results was also found in the Deary-Liewald task. A notable similarity between the results of this study and the young adults version relates to the nature of the session time effect. In both versions of the study, average response latencies in the Faces and Deary-Liewald tasks were moderately faster in the afternoon than in the morning session (this pattern will be further discussed in detail later). This result was understandable with the young adult version since they were more likely to be identified as an 'evening type' than a 'morning type'. However, I was surprised to repeat this result for the older adult version since I mainly recruited 'morning' and 'neutral types' and for that reason, I expected to observe a significant session time effect but in the opposite direction.

How do these findings relate to previous studies exploring synchrony effects among older adults (see Table 4.1)? As I highlighted in the Introduction, the evidence regarding such effects is mixed, with a number of positive findings contrasting with a range of null findings. As I noted, most extant studies are potentially compromised by low statistical power, which may have elevated the rates of false positive and negative findings. Here, my study accords with previous studies which had shown no (e.g., Rabi et al., 2022a) or only limited (e.g., Knight & Mather, 2013) evidence for a behavioural synchrony effect in tasks of this type. A practical consequence of this finding is that I believe that, at least for tasks of this type, time of testing is probably not a relevant aspect of a study to be considered, when, for instance, comparing performance of younger to older adults or comparing the performance between groups of older adults.

A number of limitations of my study are acknowledged. First, all participants completed the morning session first and the afternoon session second. This design choice was made because counterbalancing the order of sessions across participants would have implied that half of the participants would have to complete this study over three rather than two days. Since my recruitment format offered me no possibility to send automated reminders to participants, I felt that counterbalancing the session times would potentially elevate attrition rates. It is possible that the lack of counterbalancing regarding session time generated the main effect of session time which emerged in all analyses of the Faces task, with faster latencies in the afternoon (second) than in the morning (first) session. Potentially this could have been due to practice effects, which themselves may have suppressed the synchrony effects which are predicted to be larger in the afternoon than in the morning for this population. Put another way, a practice effect may have counteracted the impact of being tested at one's non-optimal time ('morning types' may have performed the Faces task in the afternoon session worse than they did if they had not previously performed the task in the morning). I did attempt to implement some control over practice effects by asking participants to first complete a practice session before undertaking the experimental sessions. Nevertheless, I acknowledge that counterbalancing the order of session times would be preferable when using a within-participants design.

A further limitation of my study is that the chronotype profile of my participants differed to some extent from the one expected based on earlier studies. While I sampled a fairly equal distribution of 'morning' and 'neutral types', May and Hasher (1993, 1998) recruited around 25% neutral older types and 75% morning older types (in both theirs and my sample, 'evening' types were virtually absent). This discrepancy could be attributed to a self-selecting bias in my recruitment. For instance, I advertised my study as requiring attendance of multiple sessions at various times of the day, and some individuals, particularly those with an 'extreme' chronotype, may be unwilling to perform the tasks at certain times of the day (e.g., those with an extreme bias towards morningness may be unwilling to attend a 4pm online session). As a consequence, individuals with 'extreme' chronotypes may have been less likely to sign up, or more likely to withdraw, from my study than 'neutral types'. The larger than expected recruitment of 'neutral types' may have potentially masked synchrony effects. I attempted to address this issue by performing analyses which involved 'morning types' only and explored whether for that sub-group, inhibitory control, response suppression and task switching are more efficient in the morning than afternoon. Results showed stable performances across the day and therefore, it is unlikely that my null finding regarding synchrony effects can be attributed to my sample's chronotype profile.

Implementing a within-participants design is inevitably more demanding for participants than a between-participants design. Attending a practice session followed by two experimental sessions may be quite demanding and for this reason, my study may have only attracted older adults who are undergoing successful aging (i.e., those who are better able to adapt to unfavourable testing times than the average older adult). Indeed, the inclusion criteria included not experiencing severe sleeping and memory problems. One may conclude from the results of my study that those undergoing successful aging are not significantly impacted by synchrony effects. This particular group of older adults may be able to use their cognitive reserves to compensate for the fact that they are tested at their non-optimal time. Instead, perhaps non-behavioural measures are needed to capture the synchrony effect among successfully aging older adults. As can be seen in Table 4.1, three studies (Anderson et al., 2014; Rabi et al., 2022a; Rabi et al., 2022b) have observed this effect in neurophysiological, but not behavioural, indices of inhibition. For instance, Rabi et al. (2022a) observed findings which imply that ERP measures of response inhibition were impacted by synchrony effects. However, the researchers did not find synchrony effects in the behavioural data. Therefore, researchers could further explore whether this (i.e., a dissociation between neurophysiological and behavioural indices) is the case. Additionally, it may be of interest for future research to investigate whether synchrony effects emerge more clearly for those who show signs of less successful aging. A recent study by Rabi et al. (2022b) tested healthy adults and older adults diagnosed with Amnestic mild cognition impairment (aMCI) on the flanker task, and a synchrony effect emerged for the latter group only. Also, under the assumption that the within-participants aspect of my study only attracted older adults undergoing successful cognitive aging, then this limitation may shed light on why my results differ from studies that reported positive findings, such as May and Hasher (1998): the latter study used a between-participants design and therefore, recruitment may not have been unintentionally limited to cognitively successful older adults.

Another potential limitation is that I did not include a noon session. As alluded to earlier, the main reason for this choice is to reduce attrition rates. Whether the exclusion of a noon session is a limitation depends on whether one would expect a 'neutral type's' optimal time of day to be at noon, as, for instance, assumed by May and Hasher (2017). Under this assumption my omission of the noon session might be problematic because 'neutral types' were not provided an opportunity to perform at their optimal time of day. Hence, it could be argued that the null findings in my study could be due to a lack of a noon session. On the other hand, May and Hasher reported no significant difference in Stroop performance between the older 'neutral types' in the morning group and those in the noon group and suggested that this reflects the liberal optimal time window of older 'neutral types'. If I am to assume that older 'neutral types' possess a wider optimal time window where performance does not significantly differ in the morning and at noon, then the lack of a noon session in my study may not be a methodological concern.

4.6 Conclusion

Older adults in my study appeared not to be susceptible to synchrony effects in inhibitory control. This pattern is not in line with previous findings suggesting that the interplay between chronotype and time of testing may influence tasks which require inhibition, but it converges with a number of studies which have also failed to report synchrony effects. Hence, my findings do not support the claim that synchrony effects are strongest for tasks which heavily rely on cognitive inhibition. However, my results may only be generalisable to a specific subset of older adults, namely those who are currently undergoing successful cognitive aging.

Chapter 5: Exploring stimulus-stimulus and stimulus-response conflict via mouse tracking

This chapter is based on:
Tseng, H., & Damian, M. F. (2023). Exploring the impact of stimulus–stimulus and stimulus–response conflicts on computer mouse trajectories: continuous flow of information from stimulus encoding to response preparation to motor action. *Psychological Research*, 1-12. https://doi.org/10.1007/s00426-023-01840-w

Author contribution: All authors contributed to the conceptualisation, material preparation and data analysis of this study. HT wrote the first draft of the manuscript, both authors contributed to the final version.

5.1 Chapter aims

In this chapter, I explored how conflicting information is processed and whether the mouse tracking paradigm can be used alongside different conflict tasks. So far, it appears that mouse tracking can be used in conjunction with tasks involving stimulus-response conflict. However, little is known on whether this paradigm is an effective tool for tasks which involve pure stimulus-stimulus conflict. I will describe two experiments involving the use of an integrated arrow task, a mixture of Stroop and Simon trials which has been claimed to measure pure stimulus-stimulus and stimulus-response inhibition. The first experiment uses keypress responses and determined whether the task can sufficiently capture these two types of conflict. More critically, the second experiment adapts this task to a mouse tracking paradigm. Therefore, Experiment 2 identifies whether the impact of S-S conflict can be captured through mouse tracking. By answering this question, I will gain insight into whether the congruency effect is countered by inhibition in a more ‘staged’ or ‘continuous’ manner

5.2 Introduction

As mentioned in Chapter 1, in conflict tasks, relevant and irrelevant stimulus dimensions and the response can engage in various forms of congruency. The ‘dimensional overlap’ (DO) model by Kornblum (1994) systematically organises various experimental tasks into a taxonomy of ‘ensembles’ depending on their dimensional overlap, with the Simon task a member of pure stimulus-response (S-R, or ‘Ensemble 3’) tasks. However, congruency effects can also arise from instances in which the response dimension is not involved, but instead relevant and irrelevant stimulus dimensions overlap. An instance of a stimulus-stimulus (S-S, or ‘Ensemble 4’ type) task is the classic colour-word Stroop task with verbal responses, in which the conflict in the ‘incongruent’ condition (e.g., naming the colour of the word BLUE printed in red) arises from two overlapping stimulus dimensions. It is worth highlighting that even in tasks characterised as involving ‘stimulus-stimulus’ conflicts here and elsewhere, it is still the case that responses might compete against each other. Even in the Stroop task, co-activated words compete with each other in the mental lexicon (e.g., Roelofs, 2003). Hence, any task which involves some sort of decision can be characterised as having a ‘response dimension’. Here, I use the term stimulus-stimulus conflict in the narrower sense that relevant and irrelevant dimensions overlap regarding their stimulus but not their response characteristics (Kornblum, 1994). This issue will be unfolded in greater detail in the General discussion section of this chapter.

Kornblum's classification of stimulus/response compatibility constellations has had a considerable impact because it relates experimental tasks to theoretical notions of perceptual and cognitive processing. His DO model (Kornblum, 1994) joins a tradition of models prevalent at the time (e.g., Sternberg, 1969) according to which information processing consists of a sequence of stages, with strict thresholding in between them. Hence, information processing at a particular stage is not initiated until processing at the previous stage has been completed to a point where a critical threshold has been reached. For instance, in Kornblum et al.'s (1999) computational implementation of the DO model, processing is divided between an input layer (concerned with stimulus encoding) and an output layer (in charge of response preparation), with a thresholding mechanism in between the two stages which gives the model serial processing characteristics. As a theoretical alternative to such strictly serial models, information may continuously 'leak' (or cascade) from one stage to the next, a notion which is embedded in numerous frameworks arising from the 'parallel distributed processing' framework (e.g., McClelland, 1979). In recent years, these continuous models have been favoured over thresholding models.

5.2.1 The characteristics of S-S and S-R effects

Over the last few decades, various empirical approaches have been developed to explore whether S-S and S-R conflict are resolved at different stages. According to the DO model, the former is processed in the stimulus encoding stage while the latter is processed in the response preparation stage. If I am to assume that this is the case, then I should expect to observe distinct characteristics of each conflict type. Indeed, there has been evidence to indicate that the S-S and S-R effects are associated with contrasting temporal dynamics. For example, Kornblum (1994) found that the S-R effect was larger when the relevant and irrelevant stimulus were presented simultaneously than when the irrelevant stimulus was displayed first. The reverse result was observed for S-S effects and hence, this difference indicates that S-S and S-R processing undergo contrasting time courses.

However, one could argue that demonstrating contrasting temporal dynamics between S-S and S-R conflicts is not sufficient evidence to suggest that these conflicts are resolved at different stages. Other strategies to distinguish between S-S and S-R conflict include factorially crossing the various types of compatibility manipulations and correlating the conflict scores with one another. The main assumption of the first strategy is that findings of additivity would indicate the two conflict types are resolved at different stages while interactivity implies that they are addressed within the same stage. Overall, the results relating to this strategy have been mixed. Some studies indicate strict additivity of both types of conflicts (e.g., Hommel, 1997, Exp. 1) whereas others exhibit underadditivity (e.g., Hommel, 1997, Exp.2). This inconsistency may be partially or wholly attributed to whether or not the ‘base’ S-S and S-R effects are substantial enough (Hommel, 1997; Rey-Mermet, 2020; Sanders, 1980). Another strategy for comparing S-S and S-R conflict is to correlate the effects produced by these conflicts with each other. Here, the logic is that a substantial correlation would point to a shared locus of the effect, whereas separate loci might predict independence. To my knowledge, only two studies have used this strategy with ‘pure’ S-S and S-R tasks (Li et al., 2014; Paap et al., 2019). Li et al. (2014) observed a null correlation on latencies but Paap et al. (2019) reported a significant correlation on 'efficiency scores’, i.e., a composite measure of speed and accuracy.

In summary, stimulus and response dimensions can be set up to conflict with one another in experimental tasks in a variety of ways. However, the exact relation between S-S and S-R effects regarding their processing characteristics remains elusive and difficult to pinpoint. This is the case even though current theoretical models tend to favour cascaded/continuous information flow over serial, staged processing.

5.2.2 The relationship between cognition and action

Similar issues arise when considering the relationship between cognition and action. A popular (and often tacit) assumption in cognitive psychology is that information flow between the final stage of cognitive processing, and of the motor stage, is thresholded: action is only released when cognitive processing has been completed. In fact, according to Calderon, Gevers and

Verguts (2018), most models, including continuous ones, follow this assumption. For instance, this notion is central to the family of 'diffusion decision' models (e.g., Ratcliff et al., 2016) according to which evidence for response alternatives accrues until it reaches a specific decision threshold for one of the responses. By contrast, the 'Unfolding Action Model' proposed by Calderon et al. (2018) suggests that information from the cognitive stage can continue to leak into action after a movement is initiated.

In strictly behavioural work, the keypress method has been used alongside conflict tasks to measure inhibition, with this ballistic response representing the end of a chain of processing stages involved in decision-making. With key presses as responses, arguably a response can only be executed after a decision has been completed. However, as argued by Wispinski, Gallivan and Chapman (2020), although it is acknowledged that data derived from such activities constitute an important source of information, a ballistic response '…does not reflect the vast majority of evolutionarily old and ecologically valid decisions for which the primate brain is organised […] competition occurs before and continues after movement initiation' (p. 40). Therefore, in this chapter, I will focus on studies which use 'dynamic' response methods to measure inhibition, such as 'mouse tracking' or 'reach tracking' (see Erb et al., 2021, Schoemann et al., 2020; Wirth et al., 2020, for recent overviews). Dynamic response methods usually involve participants completing a conflict task where responses are made by either moving the computer mouse cursor from the bottom of the screen to one of two response areas located in the upper left and right corners of the screen (in mouse tracking studies; e.g., Ye & Damian, 2023), or participants lifting their finger from a starting position typically below a screen, and touching one of two response locations on a projector screen (in reach tracking studies, e.g., Erb & Marcovitch, 2018). Mouse trajectories or hand movements are recorded, and different types of response measures are generated. These include initiation times, movement duration and reaction times. Critically, mouse or hand movement path curvatures can be captured in 'dynamic' response methods. Curvatures can be taken to reflect the online influence of conflict, with straight trajectories indicating no or little conflict, whereas deflections away from the correct response suggesting the presence of conflict. Therefore, a fundamental advantage of using this method over the keypress method is that researchers can capture the interplay between cognition and action and in turn, are able to explore the dynamics of decision-making (Freeman, Dale & Farmer, 2011; Schoemann et al., 2020).

On a very broad level, results from ‘dynamic’ response methods suggest considerable ‘leakage’ from cognition into action. Strictly serial processing would predict that motor action is only released once a decision has been fully completed. If so, the most likely outcome would be that motor activity in dynamic tasks is temporally variable in response to cognitive processing (i.e., prolonged decision making, perhaps as a result of a stimulus-response conflict, should slow down both response initiation and completion) but the shape of the response should be similar or identical. On the contrary, response curvatures are influenced by cognitive manipulations in a wide variety of contexts, which suggests that decision making affects response execution. This general observation is in line with contemporary cognitive theories in which mind, body and environment dynamically interact (e.g., Spivey & Dale, 2004).

One fundamental aspect of ‘dynamic’ response methods which so far has been unresolved will be explored in this chapter. Virtually all studies have involved some form of stimulus-response conflict which in the case of incongruency, evokes a powerful tendency to provide the incorrect response. For instance, the Simon task involves an overlap of the irrelevant stimulus with the response dimension (S-R conflict). The Simon effect appears in response latencies of key press experiments (e.g., Simon and Rudell, 1967) but it also emerges in the curvatures of dynamic responses generated with the computer mouse (e.g., Scherbaum et al., 2010; Scherbaum & Dshemuchadse, 2020) which could plausibly reflect the simultaneous activation of competing responses during movement execution. Hence, cognition ‘leaks’ into motor action. By contrast, I am not aware of studies which explored dynamic responses in conjunction with ‘pure’ stimulus-stimulus manipulations. Does a stimulus-stimulus conflict affect response trajectories in a similar way to a stimulus-response conflict?

Empirically, this question has not been directly investigated. In a recent overview of dynamic methods which mainly centred on the Flanker task, Erb et al. (2021) wrote: ‘An important question for future research to address concerns the extent to which initiation times and reach curvatures are impacted by stimulus-level and response-level conflict […] …none of the hand-tracking studies reviewed above directly evaluated the relative contributions of stimulus- and response-level conflict' (p. 743). The authors suggested that some effects on curvature might not be exclusively due to response congruency. For instance, Erb et al. (2016) reported a 'reach tracking' Stroop task in which participants classified three colours (red, blue or green) of words by reaching towards three corresponding response locations (bottom left, top centre, and bottom right). The locations for the distractors were either semantically-cued or non-cued by the target. For example, if the trial involved the word RED written in blue, the semantically-cued distractor would be red and the non-cued distractor would be green. Words could either be congruent with their colour (RED printed in red) or incongruent (RED printed in blue), and trials were analysed dependent on colour-word congruency, as well as on the congruency on the previous trial N-1. The authors also calculated distractor attraction scores representing the degree to which trajectories were more curved towards the semantically-cued distractors than the non-cued ones over time. When analysing the distractor attraction scores for incongruent trials where the previous trial was congruent, the authors observed more 'attraction' towards semantically-cued distractors than non-cued distractors and hence, demonstrating response-level conflict. However, this pattern of results did not occur when the previous trial was incongruent. Instead, the congruency effect on reach trajectories emerged for responses to the left or right location but not for responses to the centre location. According to the authors, this observation reflects the impact of stimulus-level conflict which is a delay in response selection and thus, indicating that reach trajectories are not exclusively affected by response conflicts.

Theoretically, the question of whether and how S-S conflict affects dynamically generated responses is informative based on the following argument. If S-S conflict were to emerge in a similar way to how S-R conflicts appear (mainly: via increased curvature toward the incorrect response on incongruent trials), this would imply that information continuously flows from stimulus encoding to response preparation, and eventually into motor action. This would argue against a 'thresholding' between stimulus encoding and response preparation, as hypothesised in Kornblum et al.'s (1999) model. It would further require the crossing of the 'cognition-action gap' highlighted by Calderon et al. (2018) according to which the release of a movement is dynamically gated based on the perceived amount of conflict during stimulus encoding and response preparation. Instead, S-S effects emerging in the curvature of dynamic response movements would support the idea that behaviour results from continuous information flow between perceptual, cognitive, and motor processes (Magnuson, 2005).

5.2.3 The present study

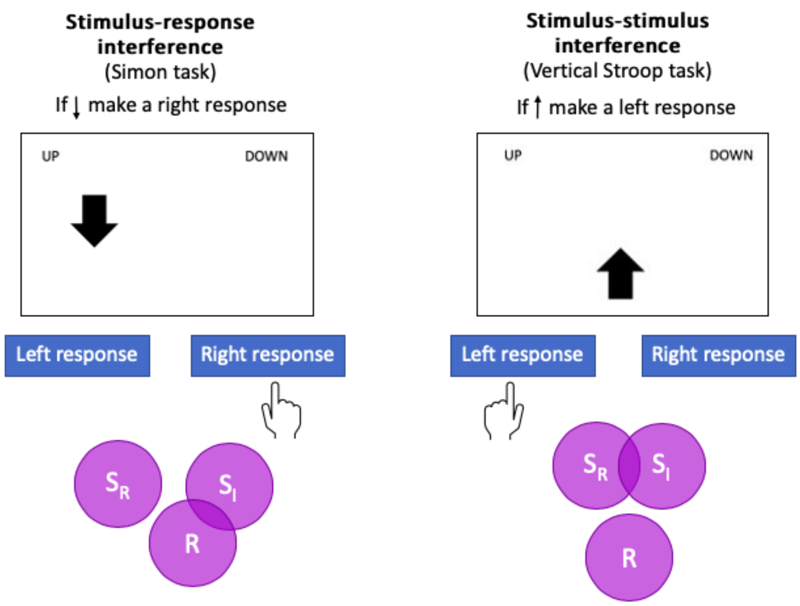
In the current study, I explored stimulus-stimulus, as well as the more standard stimulus-response based, congruency effects and their consequence on the dynamics of mouse tracking. I adopted an integrated task which generates both types of conflict within a single design and the same target stimuli, and with trials from each conflict type randomly intermixed. Figure 5.1 shows the central manipulation. The task requires participants to map up- or downward pointing arrows onto left or right responses, with the assignment of direction to response key rotated across participants. The arrows themselves can appear either to the left or right of the centre of the screen, in which case the manipulation generates a stimulus-response (in-)congruency, as in the classic Simon task (a 'type 3' conflict in Kornblum's, 1994, terminology). Alternatively, arrows can appear either above or below the screen centre (Vertical Stroop task) in which case a stimulus-stimulus (in-)congruency is generated ('type 4' conflict in Kornblum's taxonomy).

The spatial arrow task has been used in the literature before. For instance, Li et al. (2014) found S-S and S-R effects of 33 and 37 ms, respectively, with key press responses and similar effect sizes were reported by Wang et al. (2014; S-S: 30ms; S-R: 36ms). By contrast, Paap et al. (2019) reported somewhat larger effects (S-S: 77ms; S-R: 91ms) which may have arisen from the fact that a high proportion of congruent to incongruent trials (75/25) were used whereas the earlier studies used even proportions. Inclusion of a high proportion of congruent trials encourages a tonic adaptation of control strategies which results in larger conflict scores than when the proportion is low (e.g., Funes et al., 2010) but it could also induce 'contingency learning' in which participants may be able to capitalise on contingencies between stimulus and response properties (e.g., Schmidt et al., 2007). Nevertheless, these results demonstrate that with key press responses, the spatial arrow task generates powerful S-R and S-S effects.[2]

[2] See also Meier and Kane (2015) for a further study which manipulated S-S and S-R conflict type by using vertically and horizontally displaced up- and down arrows, and in which the relation between conflict resolution and individuals' working memory capacity was explored.

Figure 5.1

Experimental task. Participants classify up- or downward pointing arrows (in this case, 'if ↑ make a left response'). Stimulus-Response manipulation: arrows are presented to the left or right of screen centre. Stimulus-Stimulus manipulation: arrows are presented above or below the centre. S_R *= relevant stimulus dimension;* S_I *= irrelevant stimulus dimension; R = response dimension; correct is radiated. Both conflict types show incongruent trials. Adapted from Egner (2008) and Paap et al. (2020)*



My first experiment replicated these effects in an online task in which participants responded to the target arrows by pressing one of two response keys on the computer keyboard, operated with the index fingers of both hands. My second experiment used the same manipulation, but now in an in-person mouse tracking context in which participants initiated each trial by clicking on a 'Start' region towards the bottom of the screen, and responded by clicking on one of two 'Response' regions orientated in the top left and right corners of the screen. To prevent contingency learning (see above), I included equal proportions of congruent and incongruent trials in both studies. I predicted based on previous mouse tracking studies powerful effects of S-R conflict in errors, response latencies, and (most importantly) in the curvature of trajectories toward the incorrect response on incongruent trials. The central question was how potential effects of S-S conflict would emerge in mouse tracking, and how their dynamics would compare to those generated by S-R conflict.

5.3 Experiment 1

5.3.1 Method

5.3.1.1 Participants

30 participants were recruited online (Female = 23, Male = 7, Mean age = 24.87 years old). Participants were recruited through Prolific and received money for their participation. This experiment received ethical approval (code: 12121997085) and all participants provided their informed consent.

5.3.1.2 Materials, Design and Procedure

The study was carried out on Gorilla (http://gorilla.sc; Anwyl-Irvine et al., 2020), an online platform that enables researchers to perform psychological experiments online. As described in the Introduction, I used a modified vertical Stroop/Simon task in which participants judged the direction of upward- or downward-pointing arrows, presented either to the left or right of the fixation cross or up or down. Left- or right presented arrows formed the S-R (Simon) manipulation, and up- or down presented arrows formed the S-S (vertical Stroop) manipulation. Trials of both types were randomly intermixed. Participants were instructed to press the ‘q’ and ‘p’ keys on the computer keyboard as responses with the index fingers of their left and right hands. Assignment of up- and down-pointing arrows to response keys was counterbalanced across participants.

Participants completed 32 practice trials, followed by 240 experimental trials, with 120 Simon trials, and 120 vertical Stroop trials. Within each manipulation, 50% of trials were incongruent while 50% were congruent. In each trial, participants saw a fixation cross on a white background for 500ms in the centre of the screen, followed by a black target arrow. To remind participants of the response key assignment, the words ‘UP’ and ‘DOWN’, formatted in black ink, were located on the top left and top right of the screen. Participants were required to respond within 2000ms of target onset and if not, the next trial would immediately begin. Participants received instant feedback on whether they made a correct or incorrect response. Arrows were presented as solid black shapes on a white background (height: approx. 1° visual angle, width approx. 1.3° visual angle) and were shown to left, right, above or below the fixation dot with their centre displaced by approximately 4° visual angle (all measures are estimates because they depend to some extent on the screen on which participants performed the online experiment). An experimental session took approximately 15 minutes to complete.

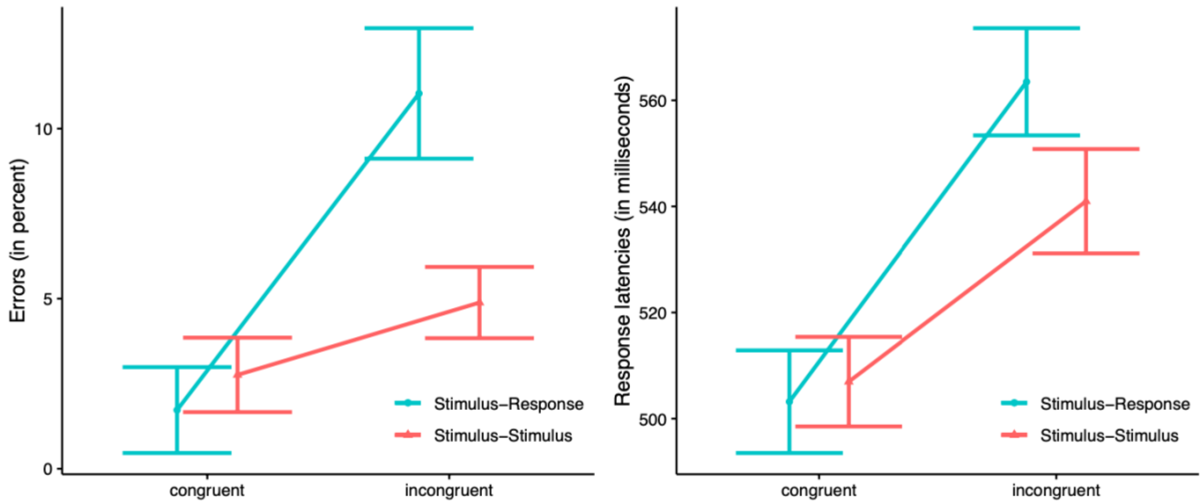
5.3.2 Results

Pre-screening revealed that one participant exhibited an error rate larger than 25%, and this participant was removed from the analysis, with 29 participants remaining. Reaction times faster than 150ms and slower than 1,500ms (1.5%), as well as latencies on error trials (4.4%), were removed from the analysis. Results are shown in Figure 5.2.

Errors were analysed via a two-way Analysis of Variance (ANOVA) with the within-participants factors conflict type (S-R vs. S-S) and congruency (congruent vs. incongruent). Results showed a highly significant main effect of conflict type, $F(1, 28) = 19.13$, $MSE = 10$, $p < .001$, of congruency, $F(1, 28) = 54.67$, $MSE = 17$, $p < .001$, and a highly significant interaction between congruency and conflict type, $F(1, 28) = 31.29$, $MSE = 12$, $p < .001$. Additionally, the Bayesian analysis revealed 'very strong' evidence in support for these effects and interaction (all Bayes factor > 100). The effect of congruency was significant for the S-R conflict type (9.3%), $t(28) = 7.50$, $p < .001$; as well as for the S-S conflict type (2.1%), $t(28) = 3.00$, $p = .005$. A parallel ANOVA conducted on the response latencies revealed a significant main effect of conflict type, $F(1, 28) = 9.33$, $MSE = 273$, $p = .005$. However, the Bayes factor for this effect was 0.51 indicating 'anecdotal' evidence supporting the null hypothesis. The ANOVA also revealed a significant congruency effect, $F(1, 28) = 110.84$, $MSE = 582$, $p < .001$, and the Bayes factor of over 100 indicates 'very strong' evidence for this effect. Finally, the ANOVA revealed a significant interaction, $F(1, 28) = 4.86$, $MSE = 1{,}030$, $p = .036$, and the Bayes factor of 23.92 lends strong support for this significant interaction. The effect of congruency was significant for the S-R type (60ms), $t(28) = 7.75$, $p < .001$, as well as for the S-S type (34ms), $t(28) = 4.78$, $p < .001$.

Figure 5.2

Experiment 1. Mean errors rates (in percent; left panel) and response latencies (in milliseconds; right panel) plotted against type of conflict (Stimulus-Response vs. Stimulus-Stimulus) and congruency (congruent vs. incongruent). Error bars reflect 95% within-participants confidence intervals (Morey, 2008)



5.3.1 Discussion

This study exhibited highly significant S-R and S-S congruency effects both in response latencies and error rates. Regarding latencies, the S-S effect (34ms) reported in my study compares well with previous studies using a similar arrow task, for example, Li et al. (2014; S-S: 33ms; S-R: 37ms) and Wang et al. (2014; S-S: 30ms; S-R: 36ms). However, in comparison to these studies, the S-R effect (60ms) in my study is noticeably larger. The reason for this discrepancy is presently unclear, nonetheless, I reliably captured S-S and S-R congruency effects in key press responses in the arrow task. In the next experiment, I tackle the central question of how these effects emerge in mouse tracking. As outlined in the Introduction, numerous mouse tracking studies have used S-R manipulations and observed more curved response trajectories for incongruent than congruent trials. The central question is whether this is also the case for S-S conflicts which are generated via conflict between relevant and irrelevant stimulus dimensions but do not involve the response dimension.

5.4 Experiment 2

5.4.1 Method

5.4.1.1 Participants

28 University of Bristol undergraduate students (Female = 21, Male = 6, Non-Binary = 1, average age = 19.71 years old) were recruited. They were rewarded with experimental credits for their participation and were tested individually in a quiet laboratory setting. This experiment received ethical approval (code: 12121997085) and all participants provided their informed consent.

5.4.1.2 Materials, Design and Procedure

This experiment replicated the online study reported as Experiment 1 as closely as possible in terms of experimental design and procedure. However, *MouseTracker* software, developed by Freeman and Ambady (2010), was used to run the task. Participants were seated approximately 60 cm from a computer screen (23 inch Dell P2319H flat screen monitor with screen resolution 1920 × 1080). Participants initiated a trial by clicking on a grey box (192 × 108 pixels; 5.1 × 2.9 cm) in the bottom centre of the screen. A fixation cross appeared in the centre of the screen for 500ms. Afterwards, the arrows were presented as solid black shapes on a white background (170 × 130 pixels; 4.5 x 3.4 cm) and were shown to left, right, above or below the fixation dot with their centre displaced by approximately 260 pixels (6.9 cm). Participants made a response by clicking on one of two response fields (288 × 144 pixels; 7.6 × 3.8 cm) on either the top left or the top right corner of the screen. Response boxes were labelled with 'UP' or 'DOWN' to indicate the corresponding response. MouseTracker recorded x and y coordinates of the trajectory of the mouse movement every 16ms for the duration of the trial (2,500ms). Participants were not able to move their cursor before the arrow was displayed. This was to prevent recording negative initiation times which would be difficult to interpret. Additionally, to ensure that I am measuring online processing and hence, participants are not using the strategy of fully committing to a decision before moving the mouse, I implemented an initiation time deadline of 250ms. Cursor speed was set in MouseTracker to a value of 12 (with 1 as the slowest and 20 as the fastest setting).

5.4.2 Results

5.4.2.1 Pre-processing

Data were processed in *R* (R Core team, 2021) using the package *mousetrap* (Kieslich, Henninger, Wulf et al., 2019). For each trial, initiation time was calculated as the time at which participants began moving the mouse, calculated relative to the onset of the target arrow (in ms). Response latencies were determined as the time between target onset and the time at which participants clicked on one of the response boxes (in ms). Errors were calculated as responses toward the incorrect response region. As a measure of the curvature of a movement trajectory, I used ‘Area under curve’ (AUC), the geometric area between the trajectory and a straight line proceeding directly from the start to the response region, calculated in pixels (Kieslich, Wulff, Henninger & Haslbeck, 2017). To avoid very large numbers, these were subsequently divided by 1,000. I would like to note that 'Maximum absolute deviation' (MAD) can also be used as a measure of mouse trajectories. This measure refers to the maximum perpendicular distance between the observed and straight trajectories. However, as can be seen in Figures 5.3 and 5.4, bimodality appears to be more clearly present in the MAD measures than in AUC ones. This was confirmed by the Hartigans' dip test for unimodality (Hartigan & Hartigan, 1985) where the statistics were significant for the MAD measures across all conditions (all Ds > .016 and ps < .006) but not for the AUC measures (all Ds < .011 and ps > .283). Considering the fact that the distribution of the MAD is bimodal, the AUC appears to be the more appropriate measure to use when performing the ANOVAs.

Figure 5.3

Experiment 2. Distributions of Area Under Curve (AUC) across conditions

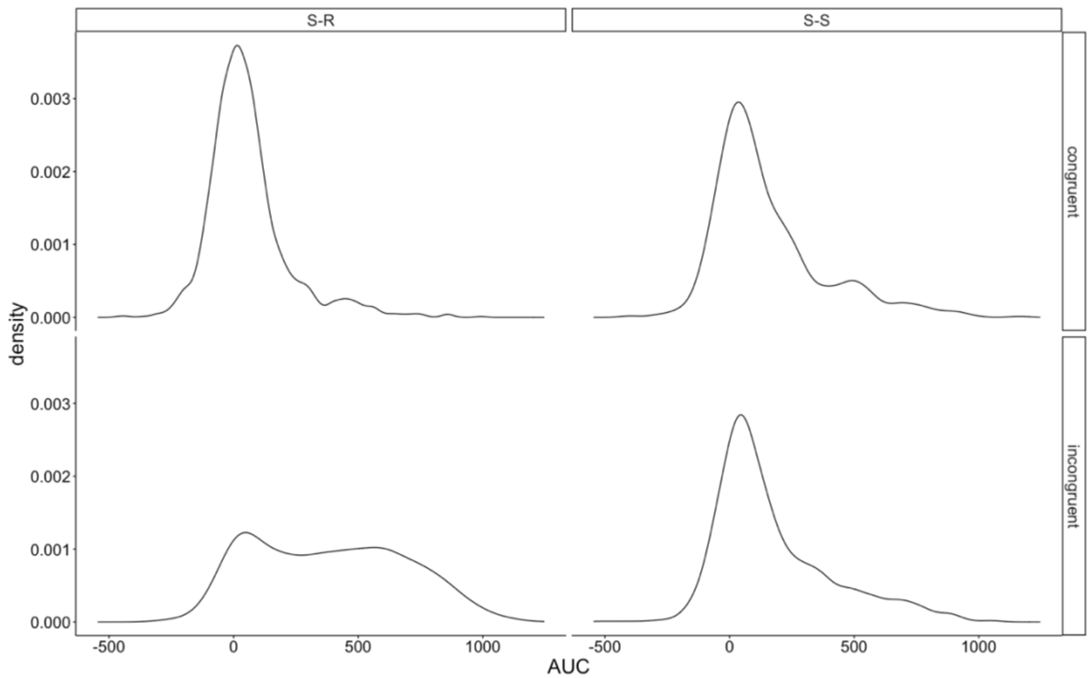
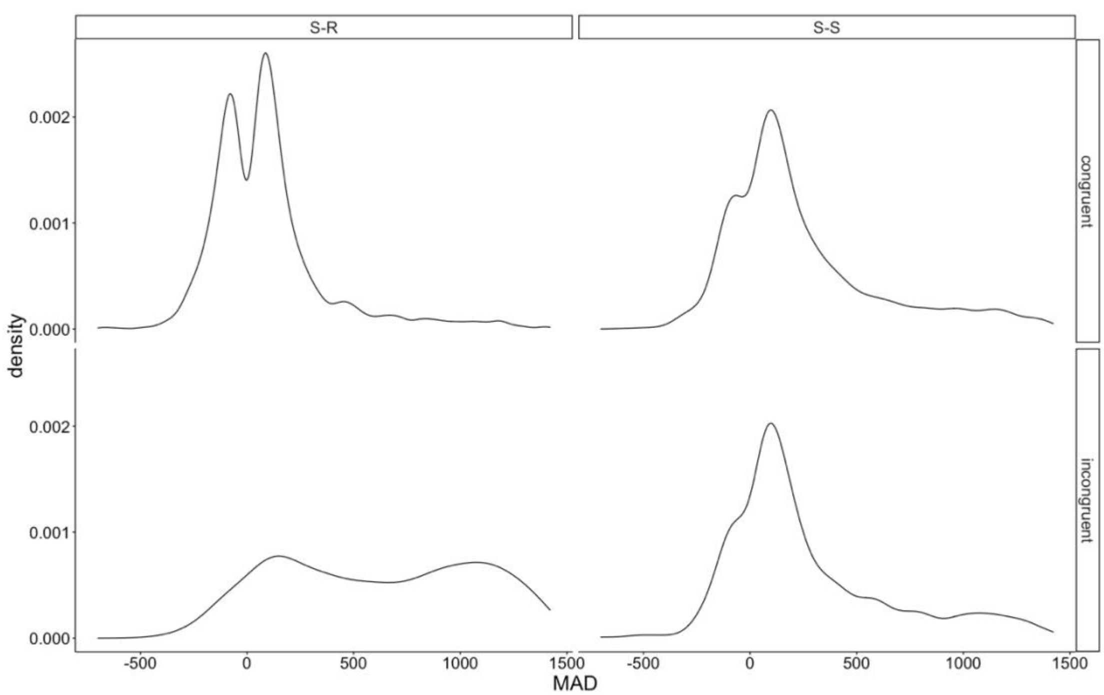


Figure 5.4

Experiment 2. Distributions of Maximum Absolute Deviation (MAD) across conditions

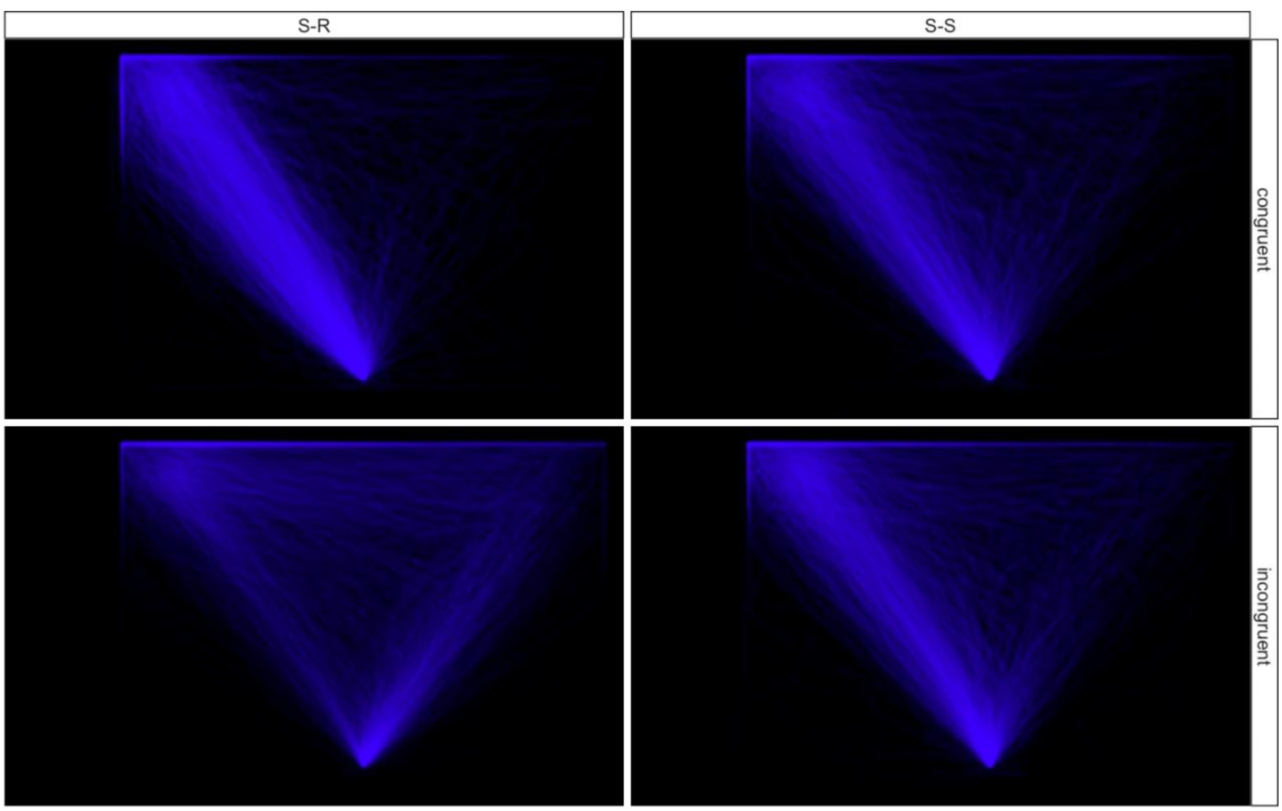


Trials with response latencies faster than 0ms and slower than 2000ms (1.6%) were excluded from the analysis, as were trials with initiation times slower than 500ms (2.6%). For the main analysis, all trajectories were time-normalised into 101 time steps and flipped to appear pointing towards the left response box. Finally, in previous experiments using MouseTracker, I observed that participants clicking on the 'Start' region often generate unintentional very small cursor movements which can be mistaken for genuine movement initiation times. Hence, in the current study, I implemented an initiation threshold such that the first time sample with a cursor movement larger than 36 pixels was taken as the initiation time. This specific value was chosen because upon clicking on 'Start', MouseTracker moves the cursor to the centre of the start region which has a height of 72 pixels. Hence, initiation was conceptualised as the time at which the cursor left a virtual circle around the starting position with a radius of 36 pixels (cf. Ye & Damian, 2023).

5.4.2.2 Final analysis

Figure 5.5

Experiment 2. Heatmaps of raw trajectories by congruency (congruent vs incongruent) and type of conflict (Stimulus-Response vs. Stimulus-Stimulus)



As can be seen in Figure 5.5, the heatmaps associated with the raw and non-time-normalised trajectories indicate visible differences in congruency conditions for the S-R trials. To demonstrate, there are clearly more straight responses in the congruent than incongruent condition. By contrast, the differences in raw trajectories are more subtle for S-S trials. The heatmaps implies that I would likely find a congruency effect in mouse trajectories for S-R trials. However, whether this would also be the case for S-S trials is uncertain.

Figure 5.6

Experiment 2. Left side: average time-normalised trajectories by congruency (congruent vs. incongruent) and type of conflict (Stimulus-Response vs. Stimulus-Stimulus). The right side shows performance in terms of errors, initiation times, response latencies, and curvature (Area Under Curve). Error bars reflect 95% within-participants confidence intervals (Morey, 2008)

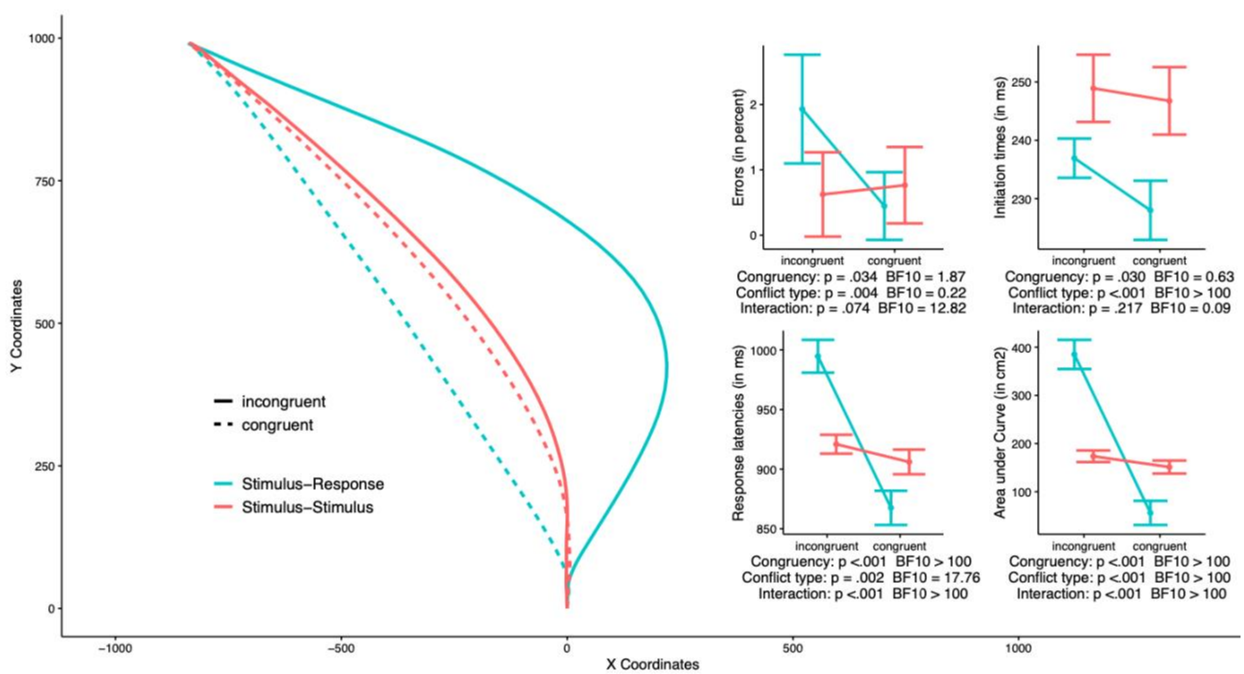


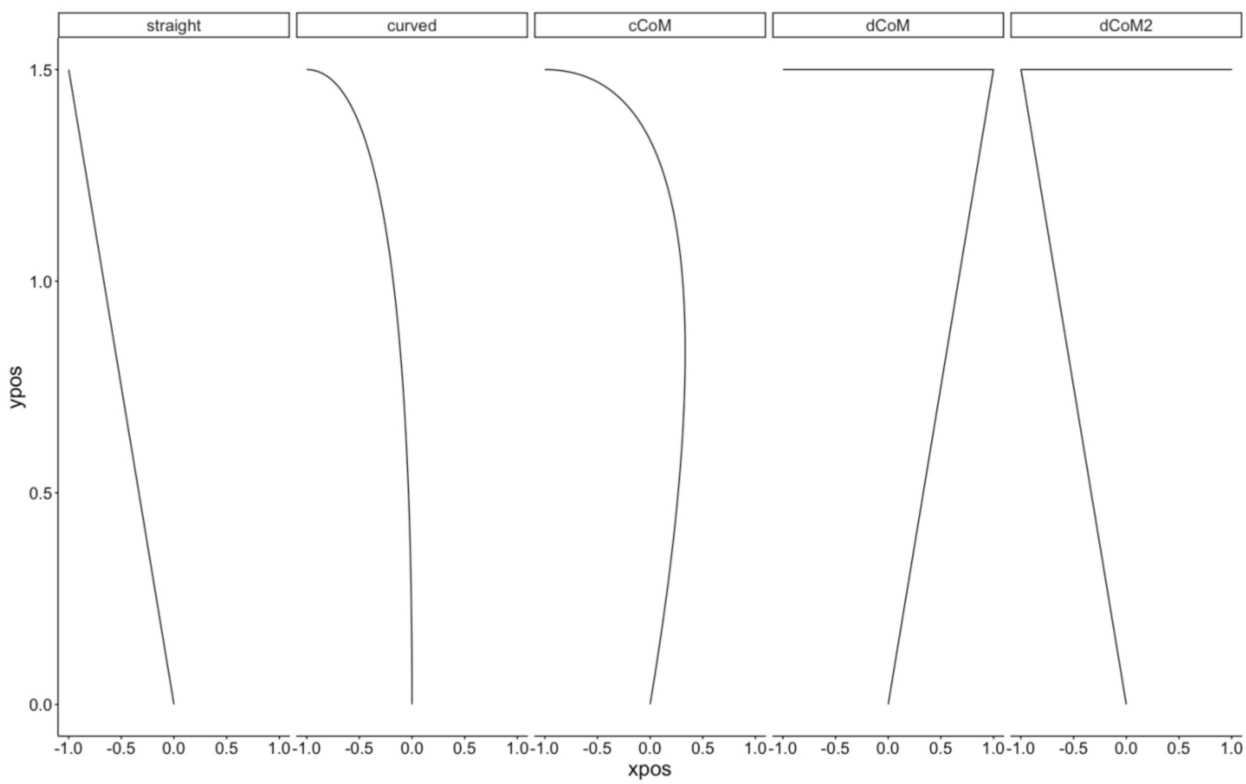
Figure 5.6 shows average mouse movement trajectories on the left side, and inset panels on the right side show the four dependent measures (error rates, initiation times, response latencies, curvature) which were analysed via two-way ANOVAs with conflict type (S-S vs. S-R) and congruency (congruent vs. incongruent). The inset panels on the right side also show the ANOVA results, with the *p* values corresponding to conflict type, congruency, and the interaction reported below each panel (detailed statistics can be found in Table A.1). Average

trajectories are less straight in the incongruent condition, when compared to the congruent condition, and this was the case not only for the S-R conflict type but also (and critically) for the S-S conflict type.

An interaction between type of conflict and congruency was found in response latencies, and AUC but not in initiation times and errors. Simple effects of congruency carried out for each conflict type separately showed that for the S-R task, the congruency effect was significant in the errors, 1.48%; $t(27) = 2.64$, $p = .014$, response latencies, 127ms; $t(27) = 11.86$, $p < .001$, initiation times, 9ms, $t(27) = 3.21$, $p = .003$, and curvature, $t(27) = 14.65$, $p < .001$. For the S-S task, congruency was not significant in the errors, -0.14%; $t(27) = 0.29$, $p = .776$, and initiation times, 2ms, t(27) = 0.50, $p = .621$, but it was significant in response latencies, 15ms; $t(27) = 2.53$, $p = .017$, and curvature, $t(27) = 2.68$, $p = .012$.

Figure 5.7

Experiment 2. An illustration of the five prototypes



As can be seen in Figure 5.6, the average trajectory corresponding to the S-R 'incongruent' condition points to the incorrect response in its early stage, then diverts course and finally arrives at the correct response. On a broad level, this 'change of mind' (CoM) indicates that the decision had not been completed when the movement was initiated but was revised during the action execution and that the incorrect response was previously activated. By contrast, for the S-S 'incongruent' condition, it is less clear whether CoM took place, and hence, whether incongruency results in an attraction towards the incorrect response. To explore this issue, I performed a 'cluster analysis' in which each raw trajectory was mapped onto one of five 'prototypes': 'straight', 'curved', 'continuous change of mind', 'discrete change of mind' and the 'double discrete change of mind' (see Figure 5.7). As you can see in Figure 5.7, two of the prototypes ('straight' and 'curved') involve no temporary diversion toward the wrong response while the three remaining ones involve various forms of CoM (see Wulff et al., 2019, for details). For that reason, I further categorised whether each raw trajectory would be classed as a CoM (i.e. 'continuous change of mind', 'discrete change of mind' and 'double discrete change of mind') or non-CoM trial ('straight' and 'curved'). An ANOVA conducted on the proportion of CoM trials showed the main effects of type of conflict, congruency, and an interaction (all *p*s $< .001$). For the S-R trials, I observed a dramatic increase of CoM trials in the incongruent relative to the congruent condition (0.07 vs. 0.55 respectively; $t(27) = 14.23$, $p < .001$); by comparison, for S-S trials the proportion of CoM trials differed less strongly and the statistical comparison was only marginally significant (0.17 vs. 0.20; $t(27) = 1.87$, $p = .072$). These results highlight the powerful attraction towards the incorrect response on S-R incongruent trials (which is exactly as expected given that the conflict involves the response dimension); by contrast, the results regarding whether incongruent S-S trials also involve activation of the incorrect response are not particularly clear.

5.5 General discussion

In my experiments, I used an integrated task previously used in the literature (e.g., Li et al., 2014) which allows the generation of stimulus-response and stimulus-stimulus compatibility effects. Experiment 1 used key press responses and confirmed that the integrated task can reliably capture both S-S and S-R effects, in line with previous studies which had used a similar manipulation (e.g., Li et al., 2014; Paap et al., 2020; Wang et al., 2014). In Experiment 2, responses were made via dynamic computer mouse movements, and the main aim was to compare and contrast the way in which S-R and S-S effects emerge in responses of this type. As highlighted in the Introduction, S-R manipulations have been widely used in conjunction with mouse tracking and typically emerge not only in response latencies but also in the curvature of the movement trajectories. The critical question was whether S-S manipulations would emerge in a similar fashion to S-R conflicts. The answer was positive: S-S effects also emerged in the curvature of response movements, although in reduced magnitude compared to S-R effects.

A methodological implication of my results is that mouse tracking can clearly be used to investigate S-S conflict. However, my findings are theoretically informative as well. As described in the Introduction, the fact that an S-R conflict consistently affects the characteristics of ‘dynamic’ responses such as mouse tracking or reaching suggests that cognition and action are not ‘staged’, i.e., motor execution begins before a decision has been fully completed. This finding is broadly in line with cognitive theories in which mind, body and environment dynamically interact (e.g., Spivey & Dale, 2004). More specifically, if there is a ‘threshold’ between cognition and action (e.g., Calderon et al., 2015) it is evidently set relatively low in my mouse tracking task such that motor activation is released at a quite early stage of the decision-making process. My novel finding is that S-S congruency affects dynamic responses. For an effect such as my current S-S manipulation (which resides at the cognitive stage of stimulus encoding) to emerge in response characteristics, it has to be assumed that stimulus encoding and response preparation are ‘cascaded’ and processing is not staged: an S-S conflict cascades into response preparation and finally emerges in motor execution. This assumption is at odds with theoretical models such as Kornblum’s (1999) DO model which proposes thresholding between stimulus encoding and response preparation. If this was true, the prediction would have been that only S-R effects appear in curvatures. This is because proponents of a model of this type would argue that an S-S conflict is resolved at the stimulus encoding stage before movement initiation while an S-R conflict would be resolved during the response preparation stage. By contrast, my findings indicate that both S-S and S-R effects can be observed in mouse trajectories. Therefore, my results are in line with ‘continuous flow’ models which bridge the gap between cognition and action (Erb et al., 2021; Eriksen & Schultz, 1979).

A potential counterpoint that can be made is that curved trajectories do not reflect online corrections in movement. For example, Song and Nakayama (2008) observed no difference in response latencies (note, the authors used the term 'total time') between curved and straight trajectories. As a result, trajectories appeared to be comparable in terms of efficiency and in turn, the authors concluded that corrective movement plans are generated before movement is initiated. This conclusion would be at odds with 'continuous flow' models which hold that movement can be executed before a decision has been completed. However, my data in Experiment 2 are not in line with Song and Nakayama's findings since it appears that curved trajectories are less efficient (indexed by response latencies) than straight trajectories. To demonstrate, in Figure 5.6 when comparing the average trajectories of S-R congruent trials with S-R incongruent trials, the former is noticeably straighter than the latter. If curved trajectories reflect corrective movement plans that were made before a participant started moving the mouse, then I should expect to observe similar response latencies between these two trial types. However, as can be seen on the response latencies graph in Figure 5.6, average response latencies are significantly longer for S-R incongruent than S-R congruent trials. Therefore, for my data, curved trajectories or larger AUCs do appear to reflect online corrections and thus, imply that participants are moving their mouse before fully committing to a decision.

It is important to embed the results of the current study in a wider literature on decision making and action. Imagine a hypothetical result of my Experiment 2 in which stimulus-stimulus incongruency resulted in a movement trajectory which compared to the congruency condition, was simply shifted in time (with slower initiation and response times) but generated trajectories of identical shape. As outlined in the Introduction, every situation which requires a decision via action involves some sort of competition between response alternatives. The hypothetical scenario would require that a) this competition took place, and was resolved, in an abstract decision space rather than between response alternatives, and that b) action commenced only after the decision was completed. Although such a scenario is not impossible and could, in fact, be predicted by 'serial' models such as those by Kornblum et al. (1999), more recent theorising tends to conceptualise decision making and action as a graded continuum. For instance, Wispinski, Gallivan and Chapman (2020) provide an extensive review and interpretation of a wealth of behavioural and neural evidence and suggest that the relationship between decision making and action is best characterised as a continuous and graded process which traverses a 'landscape' of behavioural options, from presentation until movement has been completed. This relationship can more clearly be captured using dynamic responses as opposed to ballistic ones. Under this theoretical perspective, the central finding of my second experiment which involved a dynamic rather than ballistic response method (mouse tracking), the results regarding the stimulus-stimulus compatibility condition were maybe predictable. Indeed, perhaps in this view, the surprising aspect of my findings is that stimulus-stimulus incompatibility had merely a rather subtle effect on movement trajectories. As highlighted in the Introduction, the bulk of relevant work on mouse movement trajectories has involved variations of stimulus-response compatibilities, and here, their consequences for trajectories are well-documented in numerous studies. The results of my Experiment 2 suggest that indeed, the bulk of effects on movement curvatures derives from directly competing response options, with stimulus-stimulus conflicts generating much smaller (but still reliable) consequences for trajectories. Considering the results of the change-of-mind (CoM) analysis, it is unclear whether stimulus-stimulus conflict results in an attraction towards the incorrect response or in a delay in decision making where participants move the mouse forward between the two response options for a longer time. This differs from stimulus-response conflict where it appears to be clear that conflict results in an increase in attraction towards the incorrect response.

Given that my mouse tracking experiment yielded evidence that stimulus-stimulus conflict can affect the characteristics of response movements, future research may investigate whether my findings generalise to other tasks and stimuli. For instance, in a related task used in the neuroscientific literature to disentangle stimulus- from response-based conflicts (e.g., Marinkovic, Rickenbacher & Azma, 2012), participants classify four colour patches into two responses (e.g., press 'left' if green or red; press 'right' if blue or yellow). Target colour patches are 'flanked' by two distractor patches of the same colour, which can either be congruent (e.g., a red target flanked by red distractors), stimulus incongruent (a red target flanked by green distractors), or response incongruent (a red target flanked by yellow distractors). Response latencies show a gradient, with slower stimulus incongruent than congruent responses, and slower response incongruent than stimulus incongruent responses. The prediction based on my current findings is that in a mouse tracking version of this task, average trajectories associated with the three conditions will follow the same pattern, with trajectories for the response incongruent condition showing a sizeable deflection towards the incorrect response, and trajectories for the stimulus incongruent condition still showing some deviation from the congruent condition (but probably less so than in the response incongruent condition). The same logic applies to the Eriksen task with letter flankers (Eriksen & Eriksen, 1974). Here, participants classified four target letters into two response categories (H and K, or S and C). Flanking distractor letters could be stimulus-incongruent (e.g., KKKKHKKK) or response-incongruent (SSSHSSS), and relative to a condition in which targets and flankers were identical, latencies showed a gradient with somewhat slower latencies for stimulus-incongruent flankers, and substantially slower latencies for response-incongruent flankers. Again, I predict that in a mouse tracking study, the same gradient should appear in the curvature of average response trajectories.

As summarised in the Introduction of this chapter, Li et al. (2014) factorially crossed the S-R and S-S manipulation with the same spatial arrow task used in the current study and found strict additivity in the conflict scores. A potential further experiment would be to implement this manipulation in a mouse tracking task; my prediction, based on my findings, is that interactivity rather than additivity would be observed in curvatures. This is because I found both S-S and S-R effects in curvatures, with the inference that stimulus encoding and response preparation are closely related and both cascade into motor execution. Having said that, considering that the magnitude of the S-S effects in response latencies for Experiment 2 (15ms) was relatively small, it is possible that no interaction would be found in latencies and/or curvatures because according to Hommel (1997) and Sanders (1980) interactions require relatively large base effects to be reliably observed.

5.6 Conclusion

In conclusion, my current study is to my knowledge the first to demonstrate ‘pure’ S-S effects emerging in the curvature of response trajectories in a mouse tracking task. A major theoretical implication is that stimulus encoding and response preparation are closely related (i.e., processing of the former ‘cascades’ into the latter) and both cross the gap between cognition and action (i.e., characteristics of decision making emerge in motor characteristics). A methodological implication is that mouse tracking could potentially be used in conjunction with pure S-S conflict tasks. Future research should use the mouse tracking paradigm to further investigate cognitive control via the exploration of S-S and S-R manipulations.

Chapter 6: Bilingual advantage in inhibition

Note: Experiment 2 of this chapter was pre-registered. The link to the pre-registration can be accessed below:
https://osf.io/u6nz7/?view_only=48efa6aaa88543d8a96865af1c22b526
The main aspects of the experiment that was planned include the task (i.e. the integrated arrows task) used and the protocol to continue recruiting participants until the Bayesian analyses results evoked by the group-wise comparisons provide conclusive evidence (i.e. at least 'moderate') for either the null or alternate hypothesis. Additionally, as can be seen in the pre-registration, I also planned to conduct exploratory regression analyses between task performance and various linguistic and non-linguistic measures. However, at the time of writing the pre-registration, I did not plan on which linguistic and non-linguistic measures I would utilise nor did I plan how these measures would be obtained. Adding to this, I slightly deviated from the pre-registration by asking participants to complete the Faces task as well as the integrated arrows task. This formed the basis of Experiment 1 of this chapter.

6.1 Chapter aims

The main aim of this chapter is to explore whether cognitive inhibition can be trained. Arguably, one of the most studied forms of cognitive training relates to bilingualism. Despite this area being heavily researched, there is yet a definitive answer on whether managing two or more languages incurs any cognitive benefits. Likely candidates which explain why there is such disparity in findings include publication bias, use of low sample sizes and assumptions of homogeneity among the two language groups. An added interest in exploring this research area is that I gain insight into whether the inhibitory processes involved in language management are domain-specific or domain-general.

6.2 Introduction

The perception towards bilingualism has dramatically changed over the past 100 years (Bialystok et al., 2022). Previously, a common concern expressed by parents and educators was that learning two languages at the same time would require a child to exert effort that is beyond their capacity. As a consequence, it was feared that learning a second language would interfere with their first language development and in turn, stimulate the development of language or speech disorders. For that reason, parents were often encouraged to raise their children as monolingual (Borges & Lyddy, 2023). Of course, the idea that children can be disadvantaged from learning a second language has frequently been dispelled and interestingly, the pendulum has swung in favour of bilingualism. Nowadays, you would regularly encounter news articles reporting scientific studies which suggest that bilingualism can garner cognitive advantages including improved memory, strengthened creativity and resilience to dementia. However, as mentioned in Chapter 1, research has been mixed on whether bilinguals and monolinguals substantially differ in cognitive inhibition. The following section will introduce the likely reasons for the lack of answer.

6.2.1 Potential explanations for discrepancies in findings

6.2.1.1 Publication bias

The upsurge of the published studies which report null findings from 2014 onwards (Sanchez-Azanza et al., 2017) likely reflects the impact of publication bias. de Bruin, Traccani and Della Sala (2015) followed 128 abstracts from multiple conferences and determined whether these studies were published or not. The authors reported that studies which supported the bilingual advantage were more likely to be published (68%) than studies which challenged it (29%). Put another way, this result implies the presence of publication bias. Of course, as argued by Bialystok et al. (2015), we are not informed on the reasons why some of the studies used in de Bruin et al. (2015) were not published nor do we know whether the unpublished studies were submitted for publication. For instance, some studies may have been rejected for publication due to questionable research practices or significant methodological limitations. However, given the magnitude of the difference in outcomes between the two study types, it is somewhat difficult not to conclude that publication bias has at least contributed to de Bruin et al.'s (2015) results. This is especially difficult considering the recent upsurge of published studies which report null findings (Sanchez-Azanza et al., 2017).

It should be noted that the presence of publication bias in the bilingual advantage literature does not necessarily invalidate the studies which support this phenomenon (Bialystok et al., 2015). Bialystok (2020) argues that if, hypothetically, the bilingual advantage does not exist and thus, positive findings are due to type 1 error, then we should encounter more studies reporting a monolingual advantage. Indeed, reports of a monolingual advantage are rare. However, Bialystok (2020) does not acknowledge the fact that reporting a monolingual advantage could be seen as politically undesirable. As mentioned earlier, learning a second language was deemed as a disadvantage (Darcy, 1953) and thus, children from bilingual households have been deterred from acquiring two languages. Considering the previous history of bilingualism being viewed as undesirable, one could argue that researchers may be more reluctant to submit their results for publication if their findings imply a monolingual advantage rather than a bilingual advantage. On a related note, there may be a publication bias against studies which observe superior performance in monolinguals compared to bilinguals. For that reason, the lack of published studies which report a monolingual advantage cannot be used as substantial evidence supporting the presence of a bilingual advantage. Nevertheless, publication bias itself does not entirely explain why some studies report a significant bilingual advantage. Instead, the presence of publication bias in this literature does indicate that null findings were previously under-represented.

6.2.1.2 Ceiling performance hypothesis

Major advocates of the bilingual advantage, including Bialystok, argue that the mixed behavioural findings mostly stem from studies involving young adults (e.g., Bialystok, 2016; Bialystok et al., 2005; Ware et al., 2020). As previously mentioned in Chapters 3 and 4, cognition tends to peak in young adulthood. For that reason, finding individual differences among groups from this population, who are likely to perform at ceiling, may be inherently difficult. Therefore, it may be unsurprising that results for studies which include the young adult population are mixed. Then again, it should be noted that this ceiling performance hypothesis has been disputed by findings which suggest that mixed results are also present in older adults where the bilingual advantage would be expected to be clearly present (see review by Samuel et al., 2018).

Given the rationale of the ceiling performance hypothesis, researchers have suggested that neuro-imaging methods could be used to detect the sensitive differences between the two young language groups (Bialystok, 2017). Indeed, previous research does indicate the brain structure of young mono- and bilinguals differs. For example, using fMRI, Abutalebi et al. (2012) asked young monolinguals and bilinguals to perform a Flanker task. The researchers found that bilinguals exhibited less activity than monolinguals in the dorsal anterior cingulate cortex (ACC), a region argued to be involved in conflict monitoring (e.g., Botvinick et al., 2004) and language control (e.g., Abutalebi & Green, 2007) than monolinguals. Additionally, using the Simon task, Ansaldo, Ghazi-Saidi and Adrover-Roig (2015) reported that bilinguals and monolinguals relied on different substrates: in the incongruent condition, the findings indicated that monolinguals activated the right frontal gyrus while bilinguals activated the left inferior parietal lobule. Overall, it appears to be less controversial to state that brain structural differences exist among the two language groups than behavioural ones.

Then again, observing significant brain structural differences between the two language groups could be seen as predictable. There have been numerous observations of experience-dependent plasticity, for instance, in dancers (e.g., Hufner et al., 2011) and musicians (e.g., Gaser & Schlaug, 2003; Schlaug, 2015). One of the most well-known examples of this plasticity relates to a study performed on London taxi drivers performed by Maguire et al. (2000) in which taxi drivers were reported to have larger posterior hippocampi than a control group. Considering the relatively robust evidence for experience-dependent neural plasticity, perhaps it is inevitable that a lifelong experience of managing two languages impacts an individual's brain structure. Whether reports of structural brain differences between bilinguals and monolinguals translate to evidence supporting a meaningful bilingual advantage is questionable, however. As commented by de Bruin, Dick and Carreiras (2021), it is difficult to determine whether efficient processing is reflected by low or high activation of a given brain area. Intuitively, one may immediately assume that high activation of a brain area suggests high process efficiency. However, low brain activation could also be interpreted as a sign of efficiency: one could argue that low brain activation indicates that fewer resources are needed to perform a process. Interpreting brain structural differences becomes especially difficult when considering studies, such as Ansaldo et al. (2015), which report neurological differences without the accompaniment of any behavioural differences. Ultimately, at most, neuro-imaging studies demonstrate a bilingual effect and findings from this research area alone cannot sufficiently determine whether this effect is meaningfully advantageous (e.g. Paap, 2019).

6.2.1.3 Lack of control over confounding factors

Immigration status In the older studies of the bilingual advantage, the researchers tended to recruit bilinguals who were immigrants or descendants of immigrants and compared their performance with monolinguals who tend to be non-immigrants (e.g. Bialystok et al, 2006). Differences in immigration status could be associated with differences in lifestyles, diet and education which are factors that are likely to impact cognition. Therefore, comparing two language groups with contrasting immigration statuses makes it difficult to identify a bilingual effect.

In fact, it has previously been argued that it is difficult to disentangle the effects of bilingualism from the ‘healthy immigrant effect’ (Fuller-Thomson & Kuh, 2014). The general rationale behind the ‘healthy immigrant effect’ is that ‘healthy’ people are more likely to successfully immigrate than ‘unhealthy’ people. This is because the process of immigration is challenging. The challenges faced by immigrants include learning a new language and adapting to a new culture. Further to this, the governments of countries like the USA and Canada have generated stringent eligibility criteria for immigrants. As a consequence, individuals who have job offers or high levels of education are more likely to be eligible for immigration. Considering these challenges, one could conclude that immigrants may possess superior inhibition to their monolingual counterparts and therefore, the significant language group differences found in previous studies could be attributed to immigration status as opposed to bilingualism (Fuller-Thomson & Kuh, 2014). Indeed, Kousaie and Phillips (2012) compared monolinguals to bilinguals who were not immigrants, and reported equivalent Stroop task performance between the two groups. Additionally, prior studies found a link between immigration and cognitive decline (e.g., Hill et al., 2012). Hill et al. observed that Mexican immigrants who moved to the U.S. between the ages of 20 and 49 scored higher in the mini-mental state exam (MMSE) than U.S. born participants of Mexican origin. Findings such as these lend support to the 'healthy immigrant effect' and suggest that (some) reported instances of a 'bilingual advantage' might be accounted for by immigration status rather than linguality.

Socio-economic status (SES) A further reason for mixed and inconsistent findings relates to the confounding variable of socio-economic status (SES). Higher SES has been frequently associated with better cognitive skills including inhibition, and early studies investigating the bilingual advantage did not measure or at least report the SES of the participants (e.g., Bialystok & Martin, 2004). As a result, it is unknown whether the bilinguals and monolinguals in these studies were matched on SES or whether this factor is controlled for. In turn, SES could potentially mimic or mask the effects of bilingualism. Studies which have controlled for SES (e.g., Morton & Harper, 2007; Noble et al., 2005, von Bastian, Souza & Gade, 2016) have reported null findings regarding a bilingual advantage.

Heterogeneity of bilinguals Much of bilingual advantage research tends to relegate all bilinguals into one group. This is based on the assumption that bilinguals form a homogeneous group, which is, of course, not the case (de Bruin, 2019). Bilinguals can vary according to the level of proficiency in each of their languages, the age of second language acquisition, and how they use each language, to give a few examples. Prior research indicates that performance in conflict tasks can vary within the bilingual population (e.g., Hartanto & Yang, 2016; Singh & Mishra, 2013; Tao et al., 2011). For instance, Singh and Mishra reported that bilinguals who were highly proficient in their second language performed the Stroop task better than those whose proficiency was low.

Considering these types of findings, one could assume that language group differences may be easier to identify when comparing monolinguals with certain types of bilinguals. Previous research has observed that factors such as age of second language acquisition and language proficiency could influence whether a language group difference in inhibition can be found. To demonstrate, Luk, De Sa and Bialystok (2011) reported that superior Flanker performance among bilinguals occurred when comparing monolinguals with early bilinguals (i.e., individuals who started using two languages before the age of 10) but not when compared with late ones. These findings may offer insight into why findings regarding the bilingual advantage remain mixed: perhaps, the studies reporting null findings were more likely to use the ‘wrong’ type of bilinguals than papers reporting positive ones.

Recently, numerous articles have highlighted the need to provide more detailed information regarding the language history of the bilingual sample (de Bruin, 2019). It has become standard practice within the bilingual advantage literature to include information regarding bilinguals' language proficiency and age of acquisition. Indeed, as mentioned in the previous paragraph, these two factors have been found to impact performance in conflict tasks. However, de Bruin (2019) argues that this level of detail is not sufficient and suggests that researchers need to also report other information including language use, language switching and the contexts in which bilinguals use their languages.

Green and Abutalebi's (2013) adaptive control hypothesis (ACH) provides an explanation of how the way in which a bilingual uses their languages could impact the type of cognitive benefit they gain from bilingualism. Their hypothesis introduces three types of interactional contexts which a bilingual may encounter. In a *single-language context*, a bilingual uses one language in one environment and the other language in a different environment. For example, a bilingual may speak only Cantonese at home and English at school. In this context, bilinguals would not often language switch since the languages are used in different environments. The second interactional type is the *dual-language context* where a bilingual uses their two languages in the same environment but use them with different speakers. For example, at home, a bilingual may speak Cantonese to their father but speak English to their mother. Therefore, in theory, frequent language switching does occur in this context. However, language switching within an utterance is rare. The final interactional type is the *dense code-switching context* where a bilingual regularly switches between languages within utterances.

Green and Abutalebi (2013) argue that the context which a bilingual mainly experiences determines the benefits gained from using a second language. Bilinguals in a single-language context may benefit from enhanced goal maintenance and interference control. This is since bilinguals would have to resolve the competition between their two languages. In contrast, bilinguals in a dual-language context would have the same benefits as those in a single-language context but they may also experience enhanced abilities relating to salient cue detection, selective response inhibition, task disengagement and task engagement. This is since bilinguals in this context, may, for example, have to first detect a new arrival and be required to switch languages in order to begin a conversation with the new arrival. Switching between languages would require a bilingual to disengage from the language they are currently speaking and engage in the language they want to speak. Unlike in the single- and dual-language context, the languages in a dense-coding context are in a co-operative rather than competitive relationship. As a consequence, the authors posit that the processes related to interference control are not in demand when a bilingual is in a dense-coding context. Instead, being immersed in this context enhances opportunistic planning since it is assumed that the bilingual’s choice of words is dependent on ease of access (i.e. the word that they first think of) regardless of the language it is in.

The ACH appears to provide a convincing explanation of why there are mixed results relating to the bilingual advantage. However, research investigating this hypothesis has also been mixed. For example, on one hand, Hartanto and Yang (2016, 2020) observed lower task switch costs among dual-language context bilinguals than single-language context bilinguals. This is consistent with the ACH since the former bilingual type is more likely to engage in language switching than the latter bilingual type. On the other hand, Lai and O'Brien (2020) reported only a marginally significant association between the level of dual-language context engagement and switching cost in the verbal Stroop task. Furthermore, Kalamala et al. (2020) did not find any correlation between the intensity of dual-language context experience and performance in tasks measuring 'response inhibition' (i.e. Stroop, antisaccade and go/no-go tasks). This finding has been claimed to not be in line with the ACH. This is because the researchers expected 'response inhibition' to be used more frequently in a dual-language context than a single-language context and hence, predicted that the intensity of dual-language context experience would be associated with better 'response inhibition'. Then again, perhaps Kalamala et al.'s (2020) findings may not be surprising given that advocates of the bilingual advantage have argued that this phenomenon does not reside in 'response inhibition' (Bialystok, Craik & Luk, 2012; Luk et al., 2010**,** Martin-Rhee & Bialystok, 2008).

6.2.2 The present study

In line with the overall theme of this dissertation, the focus of the present study is on cognitive inhibition. The experiments reported below investigated whether bilingualism can affect cognitive inhibition in the non-linguistic domain. In Experiment 1, I used the Faces task (which was also featured in Chapters 2, 3 and 4) to achieve this aim. To reiterate, the Faces tasks allows the capture of three distinct facets of cognitive control (response suppression, inhibitory control, and switching) within a single integrated experimental procedure. For this study, a sizeable group of mono- and bilingual young healthy adults were recruited and tested online and performed the Faces task alongside a range of other tasks and questionnaires which captured various aspects of life and language history. In the first level of analysis, I compared task performance in a group-wise manner (i.e., between monolinguals and bilinguals). Using the Faces task allowed me to explore whether bilinguals are more likely to outperform monolinguals in measures of ‘interference control’ (i.e., I use the term ‘inhibitory control’) than ‘response inhibition’ (i.e. I use the term ‘response suppression’) as claimed by some researchers (Bialystok, Craik & Luk, 2012; Luk et al., 2010**,** Martin-Rhee & Bialystok, 2008). The general rationale here is that a bilingual is more likely to encounter conflicts caused by a cue triggering two language alternatives (for instance, talking to an English monolingual about houses and the concept of houses triggers both terms in English and Cantonese) rather than conflicts which requires them to perform an action that is opposite to what the cue is triggering (for instance, speaking in Cantonese when seeing an English monolingual). Applied to the Faces task, this notion predicts a possible bilingual advantage (smaller conflict scores in bilinguals than monolinguals) for the index representing ‘inhibitory control’ and not for the index representing ‘response suppression’. There has been evidence involving children which suggests that this prediction is founded. For instance, Martin-Rhee and Bialystok (2008) observed bilingual children outperforming monolingual children on the Simon task, which is thought to involve 'inhibitory control', but comparable performances in the day-night Stroop task, which is thought to require 'response suppression'. Adding to this, Bialystok and Viswanathan (2009) asked monolingual and bilingual children to complete the Faces task (Note: my version of the task slightly differs from theirs. See Chapter 2 for details). They observed equivalent performances in the measure of 'response suppression' between the two language groups but superior bilingual performance for 'inhibitory control'. The Faces task also captures a third component of cognitive control, namely ‘task switching’. Whether a group difference emerges in switching remains to be seen (e.g., Prior & MacWhinney, 2010, but see Paap, Myuz, Anders et al., 2017).

Considering recent calls to acknowledge the heterogeneity of bilingualism, I then implemented a series of multiple linear regression analyses performed on the bilingual group only. This allowed me to explore which specific aspects of bilingualism might influence cognitive inhibition. Aspects of bilingualism which I measured include second language proficiency, age of acquisition of second language, the tendency to switch to the first language (reflects switching to L1 due to linguistic needs), the tendency to switch to the second language (reflects switching to L2 due to linguistic needs), contextual switching and unintended language switching. As you can see, this study has a special focus on language switching, in particular, since different types of language switching may require differing cognitive demands. For instance, successful contextual language switching may require more inhibitory-related processes than unintentional language switching. This is since the former measure relates to switching according to socio-linguistic factors while the latter looks into free-flowing switching. Relating back to the ACH, arguably, bilinguals in the dual- or single-language contexts would be more likely to report frequent contextual switching than dense code switchers and vice versa for unintentional switching. In fact, there is 'tentative' evidence which imply that contextual switching is associated with better inhibitory control (Jylkkä et al., 2020).

6.3 Experiment 1

6.3.1 Method

6.3.1.1 Participants

I recruited 75 monolinguals (Female = 46, Male = 29, mean age = 24.9 years old) and 75 bilinguals (Female = 48, Male = 23, Prefer not to say = 4, mean age = 25.6 years old) via Prolific where they received a monetary reward for their participation. All participants reported to be aged between 18 and 30 and were currently residing in the UK. Monolinguals were native English speakers whereas bilinguals were not. This experiment received ethical approval (code: 12121997085) and all participants provided their informed consent.

6.3.1.2 Materials

All tasks and questionnaire were carried out on Gorilla (http://gorilla.sc; Anwyl-Irvine et al., 2020).

Faces task: The trial structure (in terms of timings, stimuli used and response mappings) and the proportion of conditions in this task was the same one used in Experiment 3 of Chapter 2, Chapter 3 and Chapter 4. In total, there were 252 experimental trials and 32 practice trials (like the version used for the critical sessions in Chapters 3 and 4).

LexTALE (Lemhofer & Broersma, 2012): This task represented an objective measure of English proficiency. In this task, participants classify letter strings into words or nonwords. There were 60 letter strings in total and 40 of them were words while the rest were not. The LexTALE score, where higher scores suggest high proficiency in English, was calculated by this formula:

((number of words correct/40*100) + (number of nonwords correct/20*100)) / 2

The score which corrects for guessing ranges from 50 (chance performance) to 100 (perfect performance).

Raven's Advanced Progressive Matrices (Raven et al., 1977): I used set 1 of this task where participants encountered 12 patterns of shapes with a missing piece in the lower right corner. Participants were instructed to find this missing piece, by selecting one of eight alternatives which matches the pattern. The score in this task represents the number of correct answers produced and the maximum score was 12.

Language and Social Background Questionnaire (LSBQ; Anderson et al., 2018): The LSBQ asks participants questions in regard to their background, including their age, sex and parental education, and linguistic history, including their self-rated proficiency, age of acquiring their second language, and language use behaviour.

Bilingual Switching Questionnaire (BSWQ; Rodriguez-Fornells et al., 2012): The main reason I used this questionnaire was because one of the few well-known questionnaires which acknowledge the multi-faceted nature of language switching. For instance, in the LSBQ, only questions regarding the frequency of switching within a single utterance with various people were asked (i.e. friends, family and on social media). By contrast, the BSWQ collects information in regard to participants' code-switching experience and the reasons why they would code-switch. This questionnaire consists of 12 items and for each item, participants encountered a statement and they rated, on a five-point scale, how frequently they performed the behaviour described. A score of 1 indicates they rarely perform the behaviour while a score of 5 suggests they perform the behaviour described very frequently. The authors suggested via an exploratory factor analysis that these 12 items reflect four underlying factors which represent different aspects of language switching, with three items loading onto each factor. Hence, I computed scores (out of 15) which represented the following aspects: *switch to L1* (e.g., 'When I cannot recall a word in L2, I tend to immediately produce it in L1'), *switch to L2* (e.g., 'Without intending to, I sometimes produce the L2 word faster when I am speaking in L1'), *contextual language switching* (e.g., 'I tend to switch languages during a conversation') and *unintended language switching* (e.g., 'It is difficult for me to control the language switches I introduce during a conversation'), with higher scores implying more frequent switching.

6.3.1.3 Procedure

Participants first completed the LSBQ and BSWQ questionnaires and then performed the LexTALE and Raven's tasks. Finally, they completed the Faces task and the study lasted approximately 30 minutes.

6.3.2 Results

6.3.2.1 Pre-processing

Participants who on the Faces task produced error rates higher than 25% (n = 12) or exhibited average response latencies slower than two standard deviations from the overall mean (n = 1) in the tasks were excluded from further analysis. The analysis consisted of 67 bilinguals and 70 monolinguals.

To determine which aspects of bilingualism could influence inhibitory control, response suppression and task switching, I performed a series of multiple regressions. The various predictors that I investigated include Raven's scores, SES score, which was indexed by parental education, LexTALE score, age of acquisition, contextual switch score, unintended switch score, switch to first language score (i.e., switch to L1) and switch to second language score (i.e. switch to L2). It should be noted that participants were excluded from the analysis if they responded ‘prefer not to say’ to at least one of the questions which measured SES. Additionally, one participant was excluded from the analysis due to their LexTALE score being 50, indicating they were performing the task at chance, and one participant was excluded for producing a Raven’s score lower than 5. Ultimately, the regression analyses of the Faces task consisted of 58 participants.

6.3.2.2 Demographic information

Table 6.1 presents information regarding the demographic characteristics of the two groups. As can be seen, there are no significant differences regarding age and SES. Monolinguals outperformed bilinguals in English proficiency as measured by LexTALE. Further, a significant difference was found between the groups on Raven’s score, with bilinguals outperforming monolinguals on this task. This difference becomes a fundamental issue if I find any significant effects or interactions involving the factor language group as it may be confounded with non-verbal IQ. Although, I would like to highlight that Raven’s score did not significantly correlate with any of the performance measures (see Table B.1 for the correlation matrix table). Nevertheless, a potential solution to the Raven’s score issue is to perform an ANCOVA with Raven’s score entered as a covariate. Then again, there have been previous arguments which suggest that ANCOVAs should not be performed when there are significant group differences in the covariate of interest (Miller & Chapman, 2001). When considering this argument, the following analyses will report on the results of the ANOVAs (the results associated with the ANCOVA can be found in Table B.2).

Table 6.1
Faces task. Demographic information of the two groups of participants (monolinguals vs. bilinguals)

	Monolinguals		Bilinguals				
Measure	n	Mean	n	mean	Difference	t	p
Age	70	24.91	67	25.91	-1.00	-1.73	.085
Raven's score	**70**	**9.09**	**67**	**10.1**	**-1.01**	**-2.77**	**.006**
SES	66	4.02	60	4.22	-0.20	-0.89	.376
LexTALE	**70**	**89.25**	**67**	**81.38**	**7.82**	**4.55**	**<.001**

Note: Socioeconomic status (SES) is calculated by the average response to questions about the parent's education level: 1 – no formal education, 2- did not graduate from high school/no GCSEs, 3 – graduated from high school/GCSEs, 4 – graduated from college/6th form, 5 – undergraduate degree, 6 - Graduate or professional degree that required additional education beyond a bachelor's degree, 7 – Doctoral degree and 8 – prefer not to say. For SES, participants who responded to questions about parental education as 'preferred not to say' were removed from analysis.

6.3.2.3 Group-wise comparisons analysis

Figure 6.1 shows the results for response latencies (left panels) and error rates (right panels), separately for each cognitive component of interest (panel A: inhibitory control, panel B: response suppression, panel C: task switching) and separately for each language group (monolingual vs. bilingual). For both response latencies and error rates, I performed ANOVAs with the between-participant factor language group (monolingual vs. bilingual). For the inhibitory control analysis, the within-participant factor was congruency (congruent vs. incongruent), for response suppression, it was response type (ipsilateral vs. contralateral) and for task switching, it was task switch (task repeat vs. task switch). For each analysis, in addition to the frequentist ANOVA I also conducted a parallel Bayesian ANOVA. The results are shown in Table 6.2.

For response latencies, the Faces task was able to capture significant congruency (32ms), response type (31ms) and task switching effects (38ms). Notably, the congruency effect is of a similar size to the one I reported in Chapter 3 (27ms). However, the other two effects are larger here (in Chapter 3, response type effect = 20ms and task switch effect = 20ms). Main effects of language group were not significant, although Bayes factors qualified these comparisons as 'inconclusive'. Critically, none of the components of interest significantly interacted with group, giving no reason to suggest a bilingual advantage and hence, suggesting that conducting ANCOVAs, with Raven's score included as a covariate, (see above in section 'Demographic information') is not necessary.

For the analysis of error rates, congruency, response type, and task switch effects also emerged clearly. While the congruency and task switching effects were in the expected direction, the response type effect took the opposite direction with more errors produced in ipsilateral (8.7%) than in contralateral trials (7.2%). This pattern mirrors the findings of Chapter 3. To ensure that the main response type effect in reaction time does not represent some form of speed-accuracy trade-off, I calculated LISAS and performed an ANOVA (see Chapter 3 for more details about how LISAS are calculated). Again, like in Chapter 3, the significant main effect of response type was revealed to still be present for the ANOVA on LISAS, $F(1, 135) = 20.07$, $MSE = 2644$, $p < .001$. In the error rates, the main effect of language group was not significant, but as for the response latencies (see above), Bayesian statistics were 'inconclusive'. Critically, interactions between congruency and group, and between response suppression and group, were not statistically significant, with 'moderate' Bayesian support for the absence of interactions. The exception was that a significant interaction was obtained between group and task switching; however, Bayesian evidence regarding this interaction was 'inconclusive'. Furthermore, this interaction is no longer significant when performing an ANCOVA with Raven's score included as a covariate, $F(1, 134) = 2.34$, $MSE = 13$, $p = .161$. Adding to this, a paired sample t-test only revealed a marginally significant difference in errors for task switch trials between the two language groups, $t(135) = -1.91$, $p = .058$, and the Bayes factor of 1.94 implies only 'anecdotal' evidence supporting this interaction.

Overall, analyses conducted on both response latencies and errors show little evidence for a 'bilingual advantage'.

Figure 6.1

Faces task. Response latencies (in milliseconds; right side) and errors rates (in percent; left side). Panel A focuses on inhibitory control, panel B focuses on response suppression and panel C focusses on task switching. Each panel shows the two language groups (monolingual vs. bilinguals). Error bars reflect between-participants variability

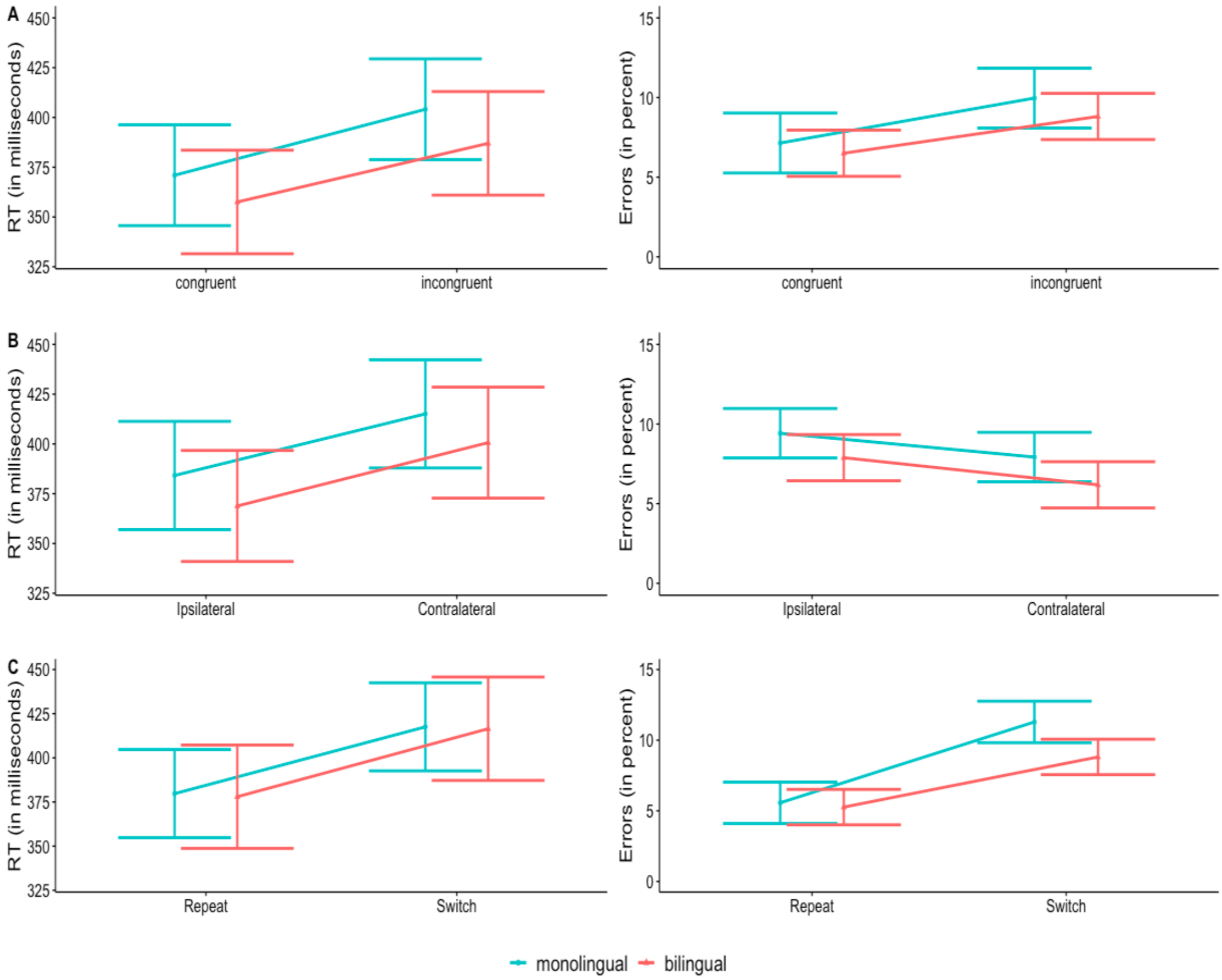


Table 6.2

Faces task. Analysis of Variance performed on response latencies and errors, separately for each component of cognitive control (inhibitory control, response suppression, task switching) and with the factor group (monolingual vs. bilingual participants). Significant results are bolded

	Response Latencies					Errors				
Effect	MSE	F	η_p^2	p	BF_{10}	MSE	F	η_p^2	p	BF_{10}
Inhibitory control										
Group	22,640	0.71	.005	.402	0.53	98	0.56	.004	.456	0.34
Congruency	**1,547**	**43.33**	**.243**	**<.001**	**>100**	**20**	**22.13**	**.141**	**<.001**	**>100**
Group × Congruency	1,547	0.15	.001	.700	0.20	20	0.22	.002	.639	0.20
Response suppression										
Group	26,057	0.58	.004	.447	0.55	78	2.35	.017	.191	0.65
Response Type	**1,494**	**45.16**	**.251**	**<.001**	**>100**	**23**	**7.60**	**.053**	**.020**	**4.63**
Group × Response Type	1,494	0.01	<.001	.926	0.16	23	0.03	<.001	.857	0.18
Task switching										
Group	24,420	0.73	.005	.590	0.62	65	2.05	.015	.154	0.64
Task switch	**878**	**115.15**	**.460**	**<.001**	**>100**	**15**	**101.11**	**.428**	**<.001**	**>100**
Group × Task Switch	878	0.06	<.001	.807	0.18	**15**	**5.55**	**.039**	**.020**	**1.94**

$df_1 = 1$; $df_2 = 135$

6.3.2.4 Regression results

As described above, following the group comparisons via ANOVAs I focussed on the bilingual group and performed multiple regressions in order to identify whether specific aspects of bilingualism could be identified which might influence cognitive inhibition. Table 6.3 shows the relevant variables and their distribution within the bilingual sample. For each cognitive component (inhibitory control; response suppression; task switching), a ‘difference score’ was computed for each participant as the dependent variable, and the predictors were entered in a simultaneous multiple regression. In addition to the frequentist regression analysis, Bayesian regressions were also performed.

The results can be seen in Tables 6.4, 6.5, and 6.6. None of the overall regression models significantly predicted performance: inhibitory control, $F(8, 49) = 0.35$, $p = .940$, adjusted $R^2 = -.100$, $BF_{10} = 0.01$, response suppression, $F(8, 49) = 1.15$, $p = 0.35$, adjusted $R^2 = .020$, $BF_{10} = 0.05$ and task switching, $F(8, 49) = 1.42$, $p = .213$, adjusted $R^2 = .055$, $BF_{10} = 0.10$. As can be seen in all of these tables, virtually none of the predictors were significant in conventional statistics (*p*s > .133); however, Bayesian statistics lend only 'anecdotal' support for these null findings. One notable finding is that L1 switch scores significantly predicted task switching scores (see Table 6.6), with ‘moderate’ Bayesian evidence supporting this finding.

Table 6.3
Faces task. Descriptive statistics of the SES scores, Raven’s scores and measures of six aspects of bilingualism (bilinguals only)

Predictor	Min	Max	Mean (SD)
SES	1	6.5	4.25 (1.30)
Raven’s score	3	12	10.10 (1.61)
Age of Acquisition (AoA)	0	26	6.81 (5.43)
LexTALE score	60	97.5	81.19 (10.78)
Contextual Switch	4	15	9.26 (2.69)
Unintended Switch	5	13	8.01 (1.48)
Switch to L1	3	13	9.03 (2.26)
Switch to L2	3	13	9.21 (2.17)

Table 6.4
Faces task. The output of the multiple regression analysis with SES scores, Raven's scores and measures of six aspects of bilingualism entered as predictors of inhibitory control score

Predictor	Estimate	SE	t	p	BF_{10}	VIF
(Intercept)	76.65	90.69	0.85	.402		
SES	5.35	6.34	0.85	.402	0.52	1.33
Raven's score	-4.97	4.62	-1.08	.287	0.67	1.09
Age of Acquisition (AoA)	-0.23	1.42	-0.16	.875	0.45	1.17
LexTALE score	-0.22	0.71	-0.32	.760	0.47	1.14
Contextual Switch	-1.09	3.56	-0.32	.761	0.46	1.81
Unintended Switch	2.31	6.12	0.38	.708	0.46	1.61
Switch to L1	2.81	3.64	0.77	.443	0.56	1.33
Switch to L2	-3.82	4.45	-0.86	.396	0.53	1.82

Table 6.5
Faces task. The output of the multiple regression analysis with SES scores, Raven's scores and measures of six aspects of bilingualism entered as predictors of response suppression score

Predictor	Estimate	SE	t	p	BF_{10}	VIF
(Intercept)	-137.92	90.25	-1.53	.133		
SES	-2.81	6.31	-0.45	.658	0.45	1.33
Raven's score	2.38	4.59	0.52	.607	0.46	1.08
Age of Acquisition (AoA)	0.80	1.42	0.56	.577	0.47	1.17
LexTALE score	0.43	0.70	0.61	.544	0.46	1.14
Contextual Switch	4.06	3.55	1.15	.258	1.07	1.81
Unintended Switch	5.94	6.09	0.98	.334	1.31	1.61
Switch to L1	4.03	3.62	1.11	.271	0.83	1.33
Switch to L2	-0.15	4.43	-0.04	.972	0.48	1.82

Table 6.6

Faces task. The output of the multiple regression analysis with SES scores, Raven's scores and measures of six aspects of bilingualism entered as predictors of task switching score. Significant results are bolded

Predictor	Estimate	SE	t	p	BF_{10}	VIF
(Intercept)	64.25	68.68	0.94	.354		
SES	-0.31	4.80	-0.06	.949	0.43	1.33
Raven's score	-1.27	3.50	-0.36	.719	0.45	1.08
Age of Acquisition (AoA)	0.23	1.08	0.23	.830	0.43	1.17
LexTALE score	-0.61	0.54	-1.13	.263	0.95	1.14
Contextual Switch	1.38	2.70	0.51	.611	0.49	1.81
Unintended Switch	-5.74	4.64	-1.24	.222	0.65	1.61
Switch to L1	**6.23**	**2.76**	**2.26**	**.028**	**4.71**	**1.33**
Switch to L2	1.69	3.37	0.501	.619	0.50	1.82

6.3.3 Discussion

In this experiment, young mono- and bilinguals performed the Faces task. To clarify, for the group-wise comparisons, a bilingual advantage would be demonstrated by a significant interaction between the cognitive component of interest (inhibitory control; response suppression; task switching) and language group. More specifically, based on a previous claim in the literature (Bialystok et al., 2012) a possible outcome consisted of a significant interaction involving the measure representing inhibitory control, but not response suppression.

I would like to first highlight that I was able to find significant congruency, response type and task switching effects for the Faces task data. This converges with the findings I reported in Chapters 2, 3 and 4 and indicates that the Faces task was able, again, to capture indices supposedly representing inhibitory control, response suppression and task switching. Fundamentally, however, I found no significant differences in indices associated with the Faces task between the two language groups (with the exception of task switching; see below). As a consequence, my study does not lend support to the claim that the bilingual advantage would emerge in measures of inhibitory control but in response suppression (e.g., Luk et al., 2010). The only significant result that emerged was an interaction between language group and task switching in errors, with bilinguals producing fewer errors on task switch trials than monolinguals. Arguably, this could be taken as evidence in support of Luk et al.'s (2010) claims when considering the fact that task switching may involve an element of inhibitory control. However, I would like to highlight that the Bayesian analysis lends only 'anecdotal' support to this significant interaction and the superior performance by bilinguals was only marginally significant. Additionally, this interaction was no longer significant when performing the ANCOVA with Raven's score included as a covariate. Therefore, when considering these caveats, I would hesitate in overinterpreting this specific result.

The regression analysis performed on the bilingual group only revealed mostly null findings indicating that many aspects of bilingualism that I measured, including age of acquisition, L2 proficiency and frequency of contextual switching, did not significantly predict indices representing inhibitory control, response suppression and task switching. One noteworthy finding was that for the task switching model, only switch to L1 emerged as a significant predictor with high switch to L1 scores being associated with high task switching costs in the Faces task. It is not clear how this association should be interpreted as this result contrasts with the previous finding by Soveri, Rodriguez-Fornells and Laine (2011) who found that higher language switching scores (note: this score was calculated by combing the switch to L1 and switch to L2 scores together) were associated with lower mixing costs in a number-letter switch task. A potential reason why I may have encountered this result is because the scores relating to switch to L1 and switch to L2 capture opportunistic language use and by extension, these scores may reflect the tendency to dense-code switch. Indeed, there were questions which asked the use of one language when access to the other language is difficult. One could argue that bilinguals who score highly in the switch to L1 and switch to L2 measures are more likely to use their languages in a dense-code switching context than those who score lowly. As mentioned earlier, the Adaptive Control Hypothesis argues that bilinguals who mainly use their language in a dense-code switching context would not benefit from enhanced cognitive control. Arguably, low switch to L1 or switch to L2 scores may indicate that the bilinguals treat their languages in a more competitive manner and by extension, their bilingual experiences incur benefits in inhibition. Therefore, the idea that switch to L1 score captures an aspect of dense-code switching could explain why the score positively correlated with the task switching score.

One informative outcome from the regression analyses is that virtually all Bayes factors were 'inconclusive', i.e., although they do not support a possible 'bilingual advantage', they also do not convincingly argue for its absence. Almost certainly the Bayesian analysis indicates that for regressions of this type, a considerable sample of individuals is needed which exceeds the one recruited in Experiment 1 (n = 58). Overall, Experiment 1 shows no evidence in favour of a bilingual advantage in cognitive inhibition with findings being mostly null for group-wise comparisons. Furthermore, my findings contradict claims that the bilingual effect would emerge more prominently for tasks relying on 'interference control' than 'response inhibition'.

A potential explanation for these null findings is that the bilingual advantage may occur in only specific forms of interference control. Indeed, Blumenfeld and Marian (2014) go one step further than Luk et al.'s (2010) claims and argue that the bilingual advantage can only be found for tasks indexing certain subcomponents of interference control, specifically those which involve S-S (stimulus-stimulus) conflict and not those which involve S-R (stimulus-response) conflict. They argued that successful management of two languages may require resolutions of S-S conflict, rather than S-R conflict. S-S conflict may need to be resolved during language comprehension and production while S-R conflict may only need to be resolved during language production. As a consequence, the bilingual advantage may be more likely to emerge for the former conflict type than the latter type. The authors compared performances elicited by monolinguals and bilinguals on the Stroop and Simon tasks. They found a larger difference in performance between the two tasks among bilinguals than monolinguals, and specifically, they observed that bilinguals performed better at the Stroop than the Simon task. The authors concluded that bilingual processing is dominated by the management of S-S conflict rather than S-R conflict. A further study by Xia et al. (2022) appears to be in line with Blumenfeld and Marian (2014); they reported language group differences in favour of bilinguals for tasks supposedly involving S-S conflict but not in tasks supposedly involving S-R conflict. Following Blumenfeld and Marian's (2014) rationale, some null findings in previous studies in the literature may have been due to implementing tasks which do not heavily rely on the ability to resolve S-S conflict. Regarding Experiment 1, a mixture of S-S and S-R conflicts is involved in the ipsilateral incongruent condition of the Faces task. When taking this into account, the null finding regarding a bilingual advantage may have arisen because the Faces task does not involve enough S-S conflict.

However, upon closer inspection of the tasks used by Blumenfeld and Marian (2014) and Xia et al. (2022), I would argue that participants' ability to resolve pure S-S conflict was not adequately operationalised. To illustrate, in Blumenfeld and Marian's version of the Stroop task, participants encountered arrows which were either pointing leftwards or rightwards and appeared to the left or right of the fixation cross. Participants were instructed to produce left responses for leftward arrows and right responses for rightward arrows. Indeed, this task involves S-S conflict since the two stimulus attributes (i.e., pointing direction and location) can conflict with each other. However, this task also contains an element of S-R conflict because the arrow location can directly contradict the response dimension. Arguably, their Stroop task could be seen as more of a Simon task and in turn, their results do not convincingly suggest that bilinguals resolve S-S conflict better than S-R conflict. In fact, a subsequent study by Paap et al. (2019) found null findings when implementing the integrated arrow task (see previous chapter), a purer measure of S-S and S-R conflict. Adding to this, the authors performed regression analyses to determine whether certain aspects of bilingual experience, including the mean number of languages used per context and frequency of switching, predict task performance. Again, they found null results.

Then again, an important methodological detail in Paap et al. (2019) that needs to be noted is that the proportions of congruent and incongruent trials used were unequal. In their study, 75% of trials were congruent and 25% were incongruent. This was to motivate high conflict scores (e.g., Funes et al., 2010) but a possible issue with this decision is that it may encourage ‘contingency learning’ (see Schmidt et al., 2007): participants may consciously or unconsciously associate aspects of the stimuli with corresponding responses, and use knowledge about the statistics of these associations to their advantage. Given this, it would be difficult to identify whether participants in Paap et al. (2019) were performing worse on incongruent trials due to the conflict present, or rather because they were not expecting an incongruent trial to appear.

Experiment 2 reported below can be seen as an extension of Paap et al.’s (2019) study. I used the same integrated arrow task that I implemented in Chapter 5 to capture ‘pure’ S-S and S-R type conflicts. The same participants as in Experiment 1 were used (in fact, data for both the Faces and the integrated arrow task were collected in the same experimental session). The main aim was to test the claim that bilinguals outperform monolinguals only on S-S conflict tasks. I used equal proportions of congruent and incongruent trials to avoid contingency learning. As in Experiment 1, the analysis first consisted of group comparisons between mono- and bilinguals, followed by multiple regressions performed on the bilingual group only, and with a special focus on whether associations can be found between language switching and S-S conflict scores.

6.4 Experiment 2

6.4.1 Method

6.4.1.1 Participants

I recruited the same participants as in Experiment 1.

6.4.1.2 Materials

The same questionnaires and tasks were used as in Experiment 1 except the main task of interest in Experiment 2 is the integrated arrow task rather than the Faces one.

Integrated arrow task: I used the keyboard version of this task as described in Chapter 5.

6.4.2 Results

6.4.2.1 Pre-processing

I employed the same exclusion criteria and analytical protocol in this experiment as I did in the previous experiment. The final integrated arrow task analysis consisted of 72 bilinguals and 74 monolinguals and the multiple regression analyses consisted of 61 bilinguals.

6.4.2.2 Demographic information

Table 6.7 displays the results of my comparisons between monolinguals and bilinguals on age, Raven's score, SES and LexTALE score. As in Experiment 1, bilinguals exhibited a significantly higher Raven's score than monolinguals and the supplementary correlation analysis (see Table B.3 for the correlation matrix table) did reveal a significant negative correlation between this factor and S-S conflict score. The group-wise comparisons reported below were also carried out as ANCOVAs with Raven's score included as a covariate. Results from these analyses can be found in Table B.4.

Table 6.7
Integrated arrow task. Demographic information of the two groups of participants (monolinguals vs. bilinguals)

	Monolinguals		Bilinguals				
Measure	n	Mean	n	Mean	Difference	t	p
Age	74	24.96	72	25.74	-0.78	-1.40	.164
Raven's score	**74**	**9.01**	**72**	**9.86**	**-0.85**	**-2.26**	**.025**
SES	70	3.93	64	4.33	-0.40	-1.85	.066
LexTALE	**74**	**89.24**	**72**	**81.41**	**7.83**	**4.68**	**<.001**

Note: SES is calculated by the average response to questions about the parent's education level: 1 – no formal education, 2- did not graduate from high school/no GCSEs, 3 – graduated from high school/GCSEs, 4 – graduated from college/6th form, 5 – undergraduate degree, 6 - Graduate or professional degree that required additional education beyond a bachelor's degree, 7 – Doctoral degree and 8 – prefer not to say. For SES, participants who responded to questions about parental education as 'preferred not to say' were removed from analysis.

6.4.2.3 Group-wise comparisons analysis

Figure 6.2 displays the response latencies (left panel) and errors (right panel) plotted against congruency and language group (monolingual vs. bilingual) and separately for each conflict type (Stimulus-Response vs. Stimulus-Stimulus). The general patterns of results, exhibiting a larger congruency effect for the stimulus-response than the stimulus-stimulus conflict in both latencies and errors, closely mirror the results found in Chapter 5. Mono- and bilingual participants appear to show largely equivalent results.

I performed ANOVAs on both reaction time and errors with the within-participant factors congruency (congruent vs. incongruent) and conflict type (S-S vs. S-R) and the between-participant factor language group (monolingual vs. bilingual). Like in Experiment 1, for each analysis, I conducted both frequentist and Bayesian ANOVAs and the outputs can be found in Table 6.8. In terms of response latencies, I observed a significant congruency effect of 40ms and a conflict type effect of 6ms with participants reacting quicker towards S-S trials than S-R ones. Adding to this, I found that these two effects interacted with each other. Simple effects analyses via paired-sample t-tests suggested that the congruency effect was significant for both S-S (29ms), t(144) = 14.74, $p < .001$, and for S-R trials (52ms), $t(144) =$ 22.41, $p <.001$. This suggests that the integrated arrow task successfully captured both the S-

S and S-R effects. The main effect of language group was not significant, but the corresponding Bayes factor indicated 'inconclusive' evidence for or against an effect of group.

More importantly, the three-way interaction between language group, congruency and conflict type, which would have been predicted if bilinguals selectively outperform monolinguals in S-S conflict trials only, was not significant with Bayesian statistics lending 'moderate' support to a null finding. The frequentist ANOVA revealed a significant interaction between language group and congruency. Figure 6.3 indicates that this interaction could stem from slightly lower response latencies for congruent trials and higher response latencies for incongruent trials among monolinguals than bilinguals. However, upon further investigation, this appears to not be the case. A paired-sample t-test suggested equivocal performance between the two language groups on congruent, $t(144) = 0.50$, $p = .617$, and incongruent trials, $t(144) = 0.22$, $p = .825$. It should be noted that the Bayesian ANOVA lends only 'anecdotal' evidence to the interaction between congruency and language group. Furthermore, in the ANCOVA with Raven's score included as a covariate (see Table B.4), this interaction was not significant, $F(1, 143) = 4.23$, $MSE = 302$, $p = .091$.

The results of the error rate analysis mostly mirror the results of the response latencies analysis with main effects of congruency (5.7%) and conflict type (2.2%) being significant and in the same direction. Additionally, I found a significant interaction between congruency and conflict type. Paired-sample t-tests implied that the congruency effect appears for both S-S (2.3%), $t(144) = 5.92$, $p < .001$, and S-R trials (9.2%), $t(144) = 14.54$, $p < .001$, in the expected direction. Finally, I did not find a significant main effect of language group nor did I observe any interactions with this factor.

Overall, the group-wise comparisons analysis reveals no convincing support for a bilingual advantage in performance on the integrated arrow tasks, and specifically, no evidence for the claim that a bilingual advantage might be specific to S-S conflict trials (but absent in S-R conflict trials).

Figure 6.2

Integrated arrow task. Mean response latencies (in milliseconds; left panel) and errors rates (in percent; right panel) and plotted against congruency (congruent vs. incongruent), conflict type (Stimulus-Stimulus vs. Stimulus-Response) and language group (monolinguals vs. bilinguals). Error bars reflect 95% between-participants confidence intervals (Morey, 2008)

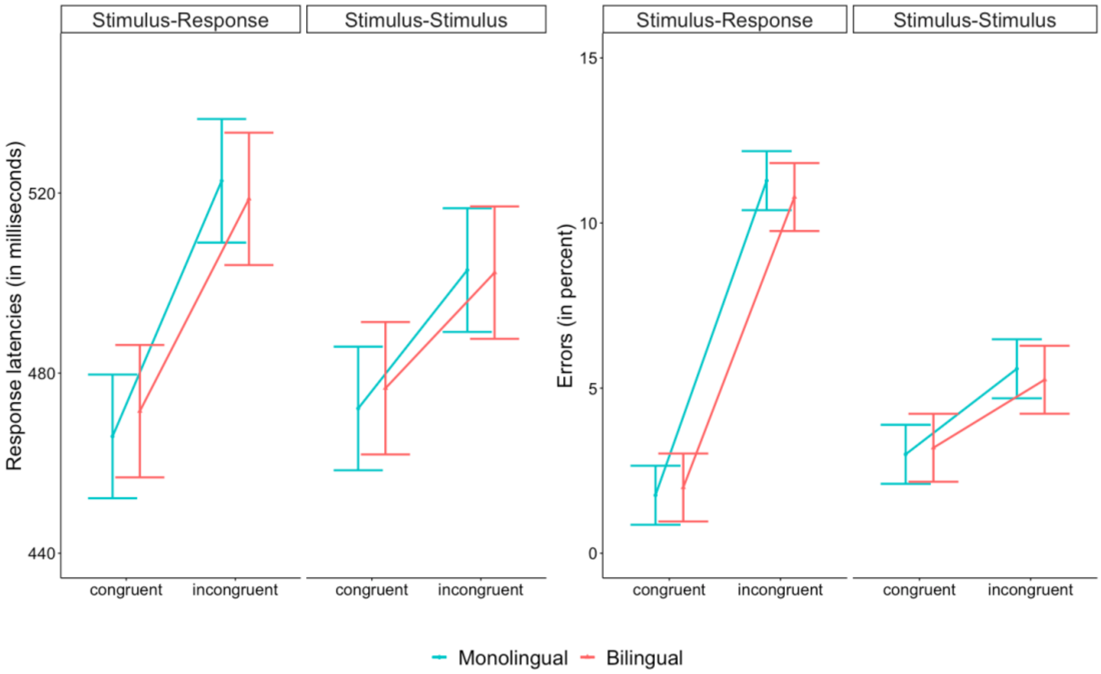


Figure 6.3

Integrated arrow task. Interaction between congruency (congruent vs. incongruent) and language group (monolinguals vs. bilinguals); see Table 6.8. Between-participants error bars were used

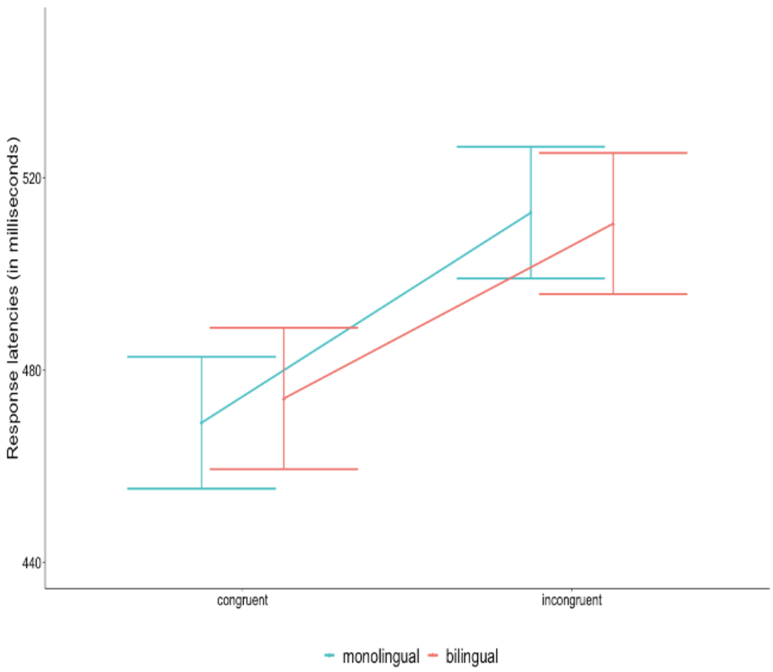


Table 6.8

Integrated arrow task. Analysis of Variance performed on response latencies and errors, with congruency (congruent vs. incongruent), conflict type (Stimulus-Stimulus vs. Stimulus-Response) and language group (monolinguals vs. bilinguals). Significant results are bolded

	Response Latencies					Errors				
Effect	MSE	F	η_p^2	p	BF_{10}	MSE	F	η_p^2	p	BF_{10}
Congruency	**312**	**752.04**	**.839**	**<.001**	**>100**	**24**	**204.25**	**.586**	**<.001**	**>100**
Conflict Type	**257**	**22.00**	**.133**	**<.001**	**>100**	**14**	**51.97**	**.265**	**<.001**	**>100**
Congruency × Conflict Type	**347**	**59.23**	**.291**	**<.001**	**>100**	**17**	**102.19**	**.415**	**<.001**	**>100**
Group	14,842	0.02	<.001	.891	0.47	68	0.02	<.001	.885	0.18
Group × Congruency	**312**	**6.29**	**.042**	**.013**	2.61	24	0.61	.004	.434	0.18
Group × Conflict Type	257	0.19	.001	.661	0.15	14	0.01	<.001	.915	0.14
Group × Congruency × Conflict Type	347	0.54	.004	.462	0.20	17	0.02	<.001	.882	0.25

$df_1 = 1$; $df_2 = 144$

6.4.2.4 Regression results

Following the group comparisons, I focused on the bilingual group and conducted a series of multiple regressions to identify whether numerous measures of the bilingual experience may influence overall congruency (computed by subtracting average response latencies for congruent trials from incongruent trials), S-S conflict and S-R conflict scores. Table 6.9 displays the relevant variables and their distribution within the bilingual sample. The predictors were entered in both a simultaneous multiple frequentist regression and a corresponding Bayesian regression.

Like for the results reported in Experiment 1, none of the overall models with SES, Raven's score, age of acquisition, LexTALE score, contextual switch score, unintentional switch score, switch to L1 score and switch to L2 score as predictors were found to be significant for overall congruency score, $F(8, 52) = 0.70$, $p = .690$, adjusted $R^2 = -.042$, $BF_{10} = 0.02$, S-R conflict score, $F(8, 52) = 0.50$, $p = .849$, adjusted $R^2 = -.071$, $BF_{10} = 0.01$, and S-S conflict score, $F(8, 52) = 0.56$, $p = .807$, adjusted $R^2 = -.063$, $BF_{10} = 0.01$. As can be seen in Tables 6.10, 6.11 and 6.12, while none of the predictors within each model reached significance (all *p*s > .149), the Bayesian analysis suggests that the evidence for these findings is 'anecdotal'.

Table 6.9
Integrated arrow task. Descriptive statistics of the SES scores, Raven's scores and measures of six aspects of bilingualism (bilinguals only)

Predictor	Min	Max	Mean (SD)
SES	1	7	4.34 (1.33)
Raven's score	5	12	10.08 (1.74)
Age of Acquisition (AoA)	0	26	6.66 (5.38)
LexTALE score	60	97	81.07 (10.76)
Contextual Switch	4	15	9.38 (2.74)
Unintended Switch	5	13	8.02 (1.40)
Switch to L1	3	13	9.08 (2.15)
Switch to L2	3	13	9.12 (2.21)

Table 6.10

Integrated arrow task – Overall congruency score. Multiple regression analysis with SES scores, Raven's scores and measures of six aspects of bilingualism entered as predictor

Predictor	Estimate	SE	t	p	BF_{10}	VIF
(Intercept)	59.88	29.45	2.03	.047		
SES	0.85	1.92	0.44	.662	0.46	1.22
Raven's score	-1.55	1.41	-1.10	.276	0.84	1.11
Age of Acquisition (AoA)	0.51	0.47	1.07	.289	0.77	1.21
LexTALE score	-0.22	0.23	-0.95	.345	0.77	1.14
Contextual Switch	0.59	1.16	0.51	.613	0.53	1.87
Unintended Switch	0.93	2.14	0.43	.666	0.51	1.65
Switch to L1	-0.36	1.24	-0.29	.772	0.46	1.32
Switch to L2	-0.95	1.37	-0.79	.489	0.53	1.69

Table 6.11

Integrated arrow task – Stimulus-Response conflict score. Multiple regression analysis with SES scores, Raven's scores and measures of six aspects of bilingualism entered as predictors

Predictor	Estimate	SE	t	p	BF_{10}	VIF
(Intercept)	91.02	48.25	1.89	.065		
SES	0.32	3.15	0.10	.919	0.45	1.22
Raven's score	-0.38	2.30	-0.17	.870	0.46	1.11
Age of Acquisition (AoA)	0.53	0.78	0.68	.501	0.62	1.21
LexTALE score	-0.40	0.38	-1.07	.292	0.74	1.14
Contextual Switch	1.03	1.90	0.54	.589	0.49	1.87
Unintended Switch	1.10	3.50	0.31	.755	0.47	1.65
Switch to L1	-1.91	2.04	-0.94	.353	0.63	1.32
Switch to L2	-1.62	2.24	-0.72	.472	0.56	1.69

Table 6.12

Integrated arrow task – Stimulus-Stimulus conflict score. Multiple regression analysis with SES scores, Raven's scores and measures of six aspects of bilingualism entered as predictors

Predictor	Estimate	SE	t	p	BF_{10}	VIF
(Intercept)	29.92	70.09	0.81	.423		
SES	1.31	2.42	0.54	.590	0.48	1.22
Raven's score	-2.59	1.77	-1.65	.149	1.06	1.11
Age of Acquisition (AoA)	0.55	0.60	0.93	.359	0.58	1.21
LexTALE score	-0.05	0.29	-0.18	.859	0.49	1.14
Contextual Switch	0.16	1.46	0.11	.914	0.48	1.87
Unintended Switch	0.74	2.69	0.28	.784	0.49	1.65
Switch to L1	1.14	1.57	0.73	.470	0.58	1.32
Switch to L2	-0.31	1.72	-0.18	.857	0.45	1.69

6.4.3 Discussion

This study tested the possibility (e.g., Blumenfeld & Marian, 2014) that bilinguals might outperform monolinguals only in tasks which implement stimulus-stimulus (SS) conflict manipulations, but not in the more generic tasks in which stimulus-response congruency is manipulated. Both S-S conflict (29ms) and S-R conflict (52ms) effects significantly emerged in the results. The magnitude of these effects is smaller than the ones reported by Paap et al. (2019; S-S effect = 75ms and S-R effect = 91ms). This is expected considering that they implemented unequal congruency proportions while I used even ones (unequal proportions are typically implemented to amplify conflict effects, but they may lead to undesirable contingency learning; see Discussion of Experiment 1). Nevertheless, this finding suggests that the integrated arrow task is a suitable tool for testing Blumenfeld and Marian's claims. To clarify, support for their claim would be demonstrated via a three-way interaction between congruency, conflict type and language group, with this interaction stemming from significant language group differences in S-S trials but not S-R ones. My results showed no evidence for such an interaction and by extension, the results do not support Blumenfeld and Marian's claims but rather my findings converge with Paap et al.'s (2019) null findings. Hence, again, I was unable to uncover any strong evidence for a bilingual advantage when comparing the performance of monolinguals to bilinguals.

As in Experiment 1, I explored whether certain aspects of bilingualism and demographical information influenced integrated arrow task performance. Through conventional statistics, I found only null results with none of the potential variables predicting any of the task indices. Again, this result is mostly in line with Paap et al.'s (2019) findings. However, surprisingly, unlike in their study, Raven's score was not found to be a significant predictor of task performance. This finding is surprising given that Raven's score has previously been observed to negatively correlate with Simon scores (e.g., Paap & Greenberg, 2013) and Stroop scores (e.g., Paap et al., 2020). Adding to this, it seems logical to predict that an association between Raven's and conflict scores would exist since the Raven's advanced progressive matrices serve to measure an individual's reasoning ability and arguably, an executive function like inhibition is important for successful reasoning. Then again, one should note that Bayesian statistics only revealed 'anecdotal' evidence in favour of this null finding and for that reason, it is difficult to arrive at any concrete conclusion. In fact, similar to the results of Experiment 1, the Bayesian analysis did not reveal anything stronger than 'anecdotal evidence' for any of the null findings.

6.5 General discussion

To recap, this study explored whether bilingualism influences and presumably improves cognitive inhibition. To achieve this, I compared the performance between young monolinguals and young bilinguals in the Faces task in Experiment 1 and the integrated arrow task in Experiment 2. By using the former task, I was able to explore whether bilingualism differentially impacts the efficiency of two types of inhibition (response suppression and inhibitory control) as well as task switching. Recall, advocates of the bilingual advantage have previously argued that bilinguals outperform monolinguals on tasks which measure interference suppression (note: I use the term ‘inhibitory control’) but not response inhibition (note: I use the term ‘response suppression’). Experiment 1 did not render evidence which supports this claim and therefore, in Experiment 2, I tested a further claim (e.g., Blumenfeld & Marian, 2014) that bilingualism might impact only certain types of interference control (stimulus-stimulus) with the integrated arrow task. Again, the evidence did not support this claim.

When putting the between-group results (both conventional and Bayesian statistics) of Experiments 1 and 2 together, there appears to be relatively strong evidence against a bilingual advantage being present, supporting a growing number of recent studies which have reported null findings (e.g., Paap et al., 2019). There are numerous implications of this finding. The results imply that the mixed findings reported in the literature are unlikely due to the question of which component of inhibition is being measured (as claimed by some researchers; e.g. Luk et al., 2010; Blumenfeld & Marian, 2014). Therefore, a theoretical conclusion that could be generated is that the inhibitory processes used for language control are domain-specific. Put another way, my study implies that far transfer effects of inhibition between the verbal and non-verbal domain do not occur.

However, this conclusion may be premature. An alternative explanation is that the bilingual advantage only occurs for specific types of bilinguals (as detailed in the Introduction). Therefore, the act of relegating bilinguals who meet the criteria required to benefit from bilingualism and those who would not into a single language group, may obscure a genuine bilingual advantage and hence, provide an explanation for my null results. I attempted to explore this issue in Experiments 1 and 2 by determining whether there are specific aspects of bilingualism that could predict task performance through regression analyses. In my study, the bilingualism variables I included were the age of acquisition, second language proficiency, and various types of language switching. Here, I assumed that if significant bilingualism-related predictors can be found, then this indicates that there are caveats to benefiting from bilingualism. For that reason, when comparing cognitive inhibition between monolingual and bilinguals, one would need to recruit specific types of bilinguals in order to uncover significant between-group results. However, my regression results from Experiments 1 and 2 did not render conclusive results. Although conventional statistics suggested that none of the potential bilingualism factors predict the scores elicited by the Faces (except for L1 switch and task switching) and integrated arrow tasks, the Bayesian analysis indicated that results were 'inconclusive' and should not be used as support for or against a bilingual advantage. The latter finding is intriguing in that it suggests that substantially larger samples than those recruited in the current study may be required to render meaningful results in regression-type explorations of individual differences among bilinguals.

Alternatively, the null findings could suggest that my assumptions regarding how the different language switching scores relate to the ACH were flawed. Recall in the introduction section, I assumed that bilinguals, who regularly use their languages within a single-language or dual-language setting, would report higher contextual switching scores. Given this assumption, I expected higher contextual switching scores to be associated with lower conflict scores. However, it could be the case that dense-code switchers, whom the ACH claims would not gain an inhibition-related benefit, also score highly in the contextual switching measure. For example, it would be unsurprising if a dense-code switcher provided a high rating in response to the contextual switching item, 'There are situations in which I always switch between the two languages'. Therefore, it is perhaps predictable that I encountered no significant relationships between contextual switching score and the various conflict scores.

Another potential explanation for the null findings relates to the use of a young sample. As mentioned in the Introduction, advocates of the bilingual advantage have argued that between-group differences are inherently more difficult to uncover with young adults (e.g., Bialystok, 2016; Bialystok et al., 2005; Ware et al., 2020) and hence, neuroimaging techniques are required to identify the subtle language group differences in this cohort. However, as previously discussed, interpreting brain structural differences should be treated with caution.

6.5.1 Limitations

A limitation of my method of comparing cognitive task performance between monolinguals and bilinguals is that I assumed that the former group was homogenous. In my study and many other studies (e.g., Hernández et al., 2010; Bialystok & Depape, 2009; Paap & Greenberg, 2013), I used monolinguals as a control group from which I identified whether bilingualism enhances inhibition. Here, I assumed that bilinguals are exercising their inhibition more than monolinguals. However, of course, heterogeneity exists not only among bilinguals but also among monolinguals as well. Some monolinguals may perform activities such as playing sports and musical instruments, which may be cognitively demanding. Indeed, there are studies which imply a cognitive benefit from undertaking these activities (e.g., Abrahan et al., 2019; Rauscher & Hinton, 2011; Granic et al., 2014). Therefore, one could argue that by not controlling for these factors, a potential bilingual advantage may be masked and that studies on this topic should capture additional information about their participants (see Paap et al., 2020, for an example of a study in which information about team sports activity, physical exercise, mediation/mindfulness, frequency of playing video games, musical training, was collected for each participant).

Likewise, a further critical assumption that I am making when comparing the two language groups is that the monolinguals used in my study are indeed purely ‘monolingual’. For instance, it is possible and in fact likely that some participants who identify themselves as monolingual have previously learnt a second language at school. In the UK, the education system encourages children and young adults to learn foreign languages. To demonstrate, for schools in England, the teaching of foreign languages in the national curriculum is required from ages 7 to 14. Furthermore, young adults are encouraged to continue learning foreign languages for their GCSEs and A-Levels. Therefore, it is questionable whether I can claim that monolinguals who were previously taught a foreign language are truly ‘monolingual’. As a result, the difference in second language knowledge between the two language groups used in my study may not be as large as I assumed. This, again, may provide insight into my null findings.

6.6 Conclusion

To summarise, for the most part, my findings suggested similar or equivalent performance between bilinguals and monolinguals, and this was regardless of which component of cognitive inhibition was measured. This indicates that the inconsistent findings within the bilingual advantage literature cannot be sufficiently attributed to the type of conflict task used or, at least, this is the case for the young adult population. While rather clear conclusions can be made based on the between-group comparisons, the picture that emerges from the regression analyses is less clear. Considering the results from the Bayesian analysis, it is questionable to make concrete inferences relating to whether aspects of bilingualism (i.e., age of acquisition, L2 proficiency and various language switching measures) or demographic measures (i.e., SES and Raven's score) can predict task performance. Therefore, although the conventional statistics indicate few significant associations between most of these predictors and various indices of cognitive inhibition, these results should be interpreted with caution. A definite conclusion regarding whether or not bilingualism conveys a genuine benefit regarding cognitive inhibition appears difficult to draw at present.

Chapter 7: General discussion

A section of this chapter parallels parts of the Discussion section of:
Tseng, H., & Damian, M. F. (2023). Exploring synchrony effects in performance on tasks involving cognitive inhibition: An online study of young adults. *Chronobiology International*. https://doi.org/10.1080/07420528.2023.2256843

7.1 Summary of findings

Recall, the main aim of this thesis was to explore the intra- (i.e., synchrony effects) and intra-individual (i.e., bilingualism) factors which influence cognitive inhibition. A somewhat secondary aim was to identify how conflict is resolved or in other words, how inhibition counters the congruency effect. In the following section, I will discuss the key findings and implications of each empirical chapter.

7.1.1 Chapter 2

This chapter acts as a precursor to the multi-part studies described in Chapters 3 and 4. An important factor that needs to be considered when designing a multi-part online study is selecting the number of tasks to use. As mentioned in the introduction section of this thesis, there are various forms of inhibition and in turn, it may be preferable to use a battery of tasks which measure these different forms. However, including too many tasks in a multi-part study may cause fatigue and elevate attrition rates. In this chapter, I examined a promising candidate, the Faces task, which has been claimed to measure three cognitive processes, inhibitory control, response suppression and task switching (Bialystok et al., 2006). Throughout this chapter, I performed a series of pilot experiments to determine whether this is the case. Unexpectedly, I found that the effects of interest were time-sensitive. To demonstrate, in Experiment 2, I introduced three different Stimulus Onset Asynchronies (SOAs) between faces cue and response asterisks and observed that the indexes for inhibitory control and task switching emerged under all the SOAs (300ms, 500ms and 700ms) while the index for response suppression emerged only at the shortest SOA (SOA 300ms). This implies that different forms/components of inhibition may undergo contrasting time courses and may be differentially sensitive to the exact timing of the manipulated task dimensions. For the last experiment, I clarified that using exclusively trials with an SOA of 300ms was able to robustly capture inhibitory control, response suppression and task switching. Given this result, I decided to use this version of the Faces task in Chapters 3, 4 and 6.

7.1.2 Chapter 3

Broadly speaking, the motivation behind this chapter (and Chapter 4) was to identify whether some of the mixed findings relating to cognitive aging and bilingualism can be attributed to the synchrony effect. In this chapter, I specifically explored potential ‘synchrony effects’ (i.e., interactions between the time of testing and individuals’ chronotypes) and recruited a large sample of young adults (over 100) to attend three experimental online sessions in the morning (8am - 9am, GMT time), noon (12pm - 1pm, GMT time) and afternoon (4pm - 5pm, GMT time). In these sessions, participants completed two tasks, the Faces (as described in Chapter 2) and the Deary-Liewald. The former was used to measure two forms of inhibition and task switching while the latter was used to measure processing speed. The results from the Faces task performance clarified that the task can capture indices representing inhibitory control, response suppression and task switching which is consistent with Chapter 2's findings. Additionally, I observed a significant time of testing effect with young adults being overall slowest in the morning sessions. However, ultimately, I did not find any support for synchrony effects on performance in this task. This result indicates that synchrony effects should not be a substantial concern when comparing cognitive task performance between groups of young adults. The chronotype profile of the sample, where the majority of young adults would be classified as ‘neutral types’, may provide insight in to this null result. Nevertheless, given the results from this chapter, I decided to not consider potential synchrony effects in Chapter 6 where I compared cognitive inhibition between young monolinguals and bilinguals. Instead, considering the significant testing time effect, I tested both language groups from noon onwards. Finally, I mentioned the idea that the synchrony effect may be identified more clearly among older adults and I empirically explored this idea in Chapter 4.

7.1.3 Chapter 4

In this chapter, I addressed whether synchrony effects are more likely to appear in older than in younger adult populations. Answering this question would especially be informative to the aging literature since recently there have been suggestions that cognitive aging may have been 'greatly exaggerated'. However, to my knowledge, most aging studies do not include information on the time of testing or participants' chronotype profile. Considering that older adults tend to be biased towards morningness and young adults tend to be biased towards eveningness, age-related differences in cognitive task performance could potentially be exaggerated or downplayed depending on the prevalent time of testing. I essentially replicated the study I performed in Chapter 3 but now with older adults. I made minor methodological changes including removing the noon session to avoid high attrition rates and adding a question about experiences with 'brain fog' associated with COVID-19. One noticeable difference I found between the older adult sample in this study and the one described in Chapter 3 is that there was less bias towards 'neutral types' in the former sample. Older adults showed somewhat slower performance compared to the younger ones tested in Chapter 4; other than that, the results relating to the older adults essentially mirrored those discussed in Chapter 3. I found this unexpected as it is generally accepted that cognitive decline occurs at older age and therefore, I predicted that older adults would be more vulnerable to synchrony effects than young adults. These results clearly imply that the synchrony effect may not have been as robust as previously thought. An alternative explanation for the null results is that my study attracted participants who are undergoing successful cognitive aging; the logic here is that my study is cognitively demanding given the multi-sessional nature of it and as a result, only those who are not particularly impacted by cognitive aging were able to complete all the sessions. When considering this rationale, it may not be surprising that I encountered null results.

7.1.4 Chapter 5

The study described in this chapter explored how the congruency effect is countered with a relatively novel paradigm, mouse tracking. A key advantage of recording responses via mouse trajectories is the wealth of information gained from it, compared to more conventional keypress responses. For instance, you can explore in more detail how conflict impacts an individual's decision making (e.g., when an individual decides to initiate a response). Previous mouse tracking studies indicate that when the information associated with stimulus and response is incompatible, participants are more likely to be swayed towards making the incorrect response. To my knowledge, most mouse tracking experiments used tasks which involve some form of stimulus-response (S-R) inhibition. Here, I investigated whether it is feasible to use the mouse tracking paradigm with a conflict task which exclusively involves stimulus-stimulus (S-S) interference. To measure pure 'S-S' and 'S-R' conflict, I used an integrated arrow task. The first experiment clarified that this task successfully captured both conflict types in keypress responses. More importantly, the second experiment with mouse tracking suggested that both conflict types emerged in curvatures of mouse trajectories. A methodological implication of this result is that the mouse tracking technique is not restricted to experimental manipulations which manipulate stimulus-response congruency, but that it is also sensitive to manipulations which generate a 'pure' stimulus-level conflict. Theoretically, the results support continuous models of perception and action which argue that information ‘leaks’ from the stimulus encoding to response preparation to motor execution stages.

7.1.5 Chapter 6

The study described in this chapter investigated the bilingual advantage. Specifically, I explored whether learning and using a second language results in a cognitive benefit among young adults. This is theoretically interesting as it provides insight into whether training inhibition in a linguistic domain can transfer to a non-linguistic domain. If this is indeed the case, a possible implication would be that inhibitory processes employed for language control are perhaps domain-general. An extensive literature has emerged on the issue of whether mono- and bilinguals differ in their inhibitory skills, with mixed results. In Experiment 1, I compared Faces task performance between young monolinguals and bilinguals. Ultimately, I observed no substantial evidence to suggest that the two language groups cognitively differ from each other. An interesting explanation for my null findings is that the bilingual advantage would only appear for certain types of conflict, in particular, stimulus-stimulus interference (Blumenfeld & Marin, 2014). Therefore, I explored whether this is the case in Experiment 2 where I asked monolinguals and bilinguals to complete the integrated arrow task (as described in Chapter 5). Considering Blumenfeld and Marin's argument, one would expect bilinguals to specifically outperform monolinguals on measures of S-S conflict but not S-R ones. However, I again encountered null results and thus, through group-wise comparison, I found no support for the claim that bilingualism incurs a cognitive benefit. For both experiments, I performed a series of regression analyses to address recent calls to explore bilingualism as a continuous factor. In this case, I implemented these analyses on the bilinguals' data and established whether there are certain aspects of bilingual language use or demographic factors which might influence various subcomponents of inhibition. Overall, the results were unclear. Through conventional statistics, the results suggested no associations between any of the predictors and scores evoked by the Faces and integrated arrow tasks. However, Bayesian analysis provides only 'anecdotal' evidence for these findings. Considering the findings elicited by the regression analyses, I was unable to draw concrete conclusions on whether inhibition is affected by bilingualism.

7.2 Implications

7.2.1 Inhibitory deficit theory

The findings of Chapters 3 and 4 are inconsistent with one of the assumptions of the inhibitory deficit model. Against the model's predictions, I was unable to demonstrate the presence of a synchrony effect in cognitive inhibition. An implication of this finding is that it indicates that the mixed findings for research regarding cognitive aging and bilingualism cannot be sufficiently explained by a lack of control over synchrony effects (as speculated in Chapter 1). Further to this, the model argues that a dissociation exists between processing speed and inhibition. However, I did not observe substantial evidence supporting this claim with time of day effects influencing performance in both the task requiring inhibition and task with minimal need of this construct. Secondly, the main findings from the young (Chapter 3) and older adults (Chapter 4) did not significantly differ from each other with neither indicating the presence of synchrony effects. This similarity in findings is not in line with the inhibitory deficit model where stronger synchrony effects among older adults than young adults would have been predicted. Indeed, one could argue that this thesis appears to be consistent with the claim that the impact of cognitive aging has been exaggerated in the literature. However, given that I did not make any direct comparisons between the two age groups in this thesis, concrete conclusions surrounding this topic cannot be made.

7.2.2 Usefulness of online testing

Except for Experiment 2 in Chapter 5, all of the studies included in this thesis were conducted online. A growing number of published web-based studies in the cognition literature (Stewart, Chandler & Paolacci, 2017) indicate that web-based studies can reproduce well-established effects including Flanker and Simon (Crump, McDonnell & Gureckis., 2013). Furthermore, the performance of participants tested online does not appear significantly different from the performance of participants tested in the lab (Casler, Bickel & Hackett, 2013; Cyr, Romero & Galin-Corini, 2021; Germine et al., 2012; Semmelmann & Weigelt, 2017). Indeed, throughout this thesis, I found that the conflict tasks were able to capture the desired effects. For example, in Experiment 1 of Chapter 5, I performed the integrated arrows task online and was able to produce the pure S-S and S-R effects like in previous studies which implemented in-person testing (e.g., Paap et al., 2019; Li et al., 2014). Therefore, the findings of this thesis are in line with studies which suggest that online testing can capture effects produced via in-person testing and hence, is a valid alternative to laboratory-based research.

The fact that online testing appears to be able to produce the same effects as in-person testing is especially promising when considering that the key advantage of conducting studies online relates to recruitment. Online testing provides researchers the opportunity to easily recruit a large sample of participants outside the university student population and thus, generate findings that are more generalisable to the general population than they would be through lab testing. For example, I was able to recruit over 100 older adults online in Chapter 4. Recruiting this number of older adults for in-person testing would have been substantially more challenging. Doing so would have been especially difficult during the COVID-19 pandemic where older adults are particularly vulnerable to COVID-19 and were advised to stay at home.

7.2.3 Usefulness of Bayesian statistics

Throughout this thesis, I used Bayesian statistics as well as conventional ones to analyse the results. While conventional statistics tends to use the *p*-value to determine whether there is enough evidence to reject the null hypothesis, results are prone to misinterpretation (e.g., Wagenmakers, Marsman, Jamil et al., 2018) and a problem with this analytical method is that it does not allow researchers to identify how strong the evidence in favour of the null compared to the alternative hypothesis. By contrast, Bayesian statistics allows us to gauge the relative evidence for or against a particular prediction. Bayesian analysis becomes especially important when examining how inhibition is influenced by synchrony effects and bilingualism, i.e., questions where previous findings have rendered inconsistent results. In my case, the Bayesian analysis suggested at least 'moderate' evidence against the presence of a synchrony effect among young adults (Chapter 3) and older adults (Chapter 4), and against a bilingual advantage (when comparing monolinguals and bilinguals; Chapter 6). However, it is worth highlighting that I sometimes found disparities between conventional and Bayesian statistics. For example, through conventional statistics, I found some effects and interactions to be significant, but with Bayesian analysis, there was only 'anecdotal' evidence supporting this conclusion. To demonstrate, in Experiment 1 of Chapter 6, I observed a significant interaction in error rate between task switching and language group. Since the nature of the interaction was in favour of the bilinguals, one could conclude that this is evidence supporting the presence of a bilingual advantage. However, the Bayesian analysis offered only 'anecdotal' evidence supporting this interaction. Therefore, I feel that researchers, especially when group-wise comparisons are involved, should consider analysing their data via both conventional and Bayesian statistics to gain a better grasp on what they can conclude from their results and to avoid over-interpretations of positive findings.

7.3 Limitations

7.3.1 Small effect sizes of the Faces task

A common theme throughout this thesis (with the exception of Chapter 5) is the use of the Faces task. Recall, the main reason I used this task was the fact that it has been argued to measure indices representing three cognitive components: response suppression, inhibitory control and task switching within a single, integrated procedure (Bialystok et al., 2006). A major advantage of using this task is that I was able to explore whether the indices representing these components are influenced by synchrony effects without the need to use multiple tasks. Adding to this, I was able to use this task to identify whether the bilingual advantage is more likely to emerge for measures of inhibitory control than response suppression as previously claimed (e.g., Bialystok, Craik & Luk, 2012; Luk et al., 2010**,** Martin-Rhee & Bialystok, 2008). In Chapters 2, 3, 4 and 6, the effects representing each component were consistently found via conventional statistics and there was at least ‘strong’ Bayesian evidence in their support. Then again, as briefly mentioned in Chapter 2, the corresponding effects were small. For that reason, arguably, the use of the Faces task in Chapters 3 and 4 made it inherently difficult to observe a significant synchrony effect in cognitive inhibition. The same reasoning can be applied to the group-wise comparisons made between monolinguals and bilinguals in Experiment 1 of Chapter 6. A larger effect size than the one made available by the Faces task would be desirable when exploring relatively subtle potential effects such as synchrony effects, and bilingualism.

Additionally, it should be noted that I did not investigate the construct validity of the Faces task. This may prove to be critical given that Rey-Mermet et al. (2018) reported very small inter-task correlations between tasks which measure inhibition. Of course, one way to address this issue is to not rely on one task to measure inhibition. Indeed, I pursued this strategy in Chapter 6 regarding the bilingual advantage. However, I exclusively used the Faces task in the chapters associated with the synchrony effect, the rationale here being that I asked participants to complete multiple sessions and hence, asking participants to complete numerous tasks multiple times may fatigue them and result in high attrition rates. Given this issue, the next logical step derived from this thesis would be to investigate the construct validity of the Faces task. This can be achieved by asking participants to complete multiple inhibition tasks and correlate them with each other. For instance, one could correlate the response suppression score captured by the Faces task with indexes captured by the stop-signal and go/no-go tasks.

7.3.2 Reliance on an online sample

Although online testing can reap many benefits, as mentioned earlier, I acknowledge that online testing comes attached with a lack of control over factors including hardware and the environment. As a result, these factors may have influenced my findings. The consequence of the lack of control over hardware is variations in timing accuracy between participants. This issue may be particularly problematic in Chapter 6 where the main analysis involved comparing performance in the Faces task and integrated arrow task between participants. For instance, participant X may appear to perform the Faces task faster than participant Y because the former completed the task with a newer laptop than the former. Then again, I would argue that variation in timing accuracy is less of a concern for studies which mainly incorporate a within-participants design. This is because one can assume that most participants would be using the same hardware to perform a study. For example, in Experiment 1 of Chapter 5, it is unlikely that participants would switch hardware during the completion of the integrated arrows task.

A factor that I was unable to address for all the online studies relates to the environment. Conventionally, in a lab setting, the researchers can ensure that participants are completing the study in a quiet environment and hence, the presence of distractions is kept at a minimum. In my online studies, I implemented screening procedures such that individuals with suspiciously high error rates were excluded from analysis, and I assumed that participants with low error rates were completing the tasks in a workable environment and were completing the study without too many distractions. This assumption is founded for studies involving the Faces task as it does require participants to dedicate their attention to the screen. To demonstrate, to produce a correct response, they must see the cartoon face and the location in which the cue appears. They would not be able to successfully perform the task without either of these pieces of information. Then again, a more robust method of determining whether a participant is working in an environment that is not too full of distractions would be to perform attention checks and exclude participants who fail them. Of course, another viable way to address this issue is to design the tasks to be more engaging so that participants would not turn to distractions.

Another potential pitfall of online studies is that the recruited participants may have completed many cognitive studies to the extent that they are non-naïve. Indeed, there have been reports of 'professional' participants who complete studies on online platforms such as MTurk as their source of income (e.g., Casey et al., 2017). A consequence of using non-naïve participants includes attenuated effect sizes (Chandler et al., 2015) and the findings generated may not be entirely generalisable. In Chapters 5 and 6, I used the integrated arrows task which involves a combination of Stroop (S-S) and Simon (S-R) trials. As mentioned in Chapter 1, both the Simon and Stroop tasks are popular tools used to measure cognitive inhibition, and for that reason, it is possible that some of the participants I recruited may not have been naïve. However, it should be noted that this issue regarding the use of non-naïve participants is not exclusive to online testing. Many in-person studies tend to recruit university students who study Psychology and are completing the study for experimental credits (e.g., May & Hasher, 2017; Paap & Greenberg, 2013). Adding to this, I would argue that this issue is somewhat circumvented for studies which involved the Faces task (Chapters 2, 3, 4 and 6). This is because when exploring the published studies which investigate cognitive inhibition, the Faces task does not appear to have been frequently used as a measure of this construct. Consequently, when signing up for an online study, participants are less likely to previously have encountered the Faces task than the Stroop or Simon tasks and in turn, are more likely to complete the task as a naïve participant.

7.3.3 Statistical power

A potential concern relates to statistical power. I did not perform any power analyses to determine the sample size I should aim to recruit. The main reason for this is because I found it difficult to identify the effect sizes I should input for the power analyses especially in my multi-part studies. Indeed, I did conduct a series of pilot studies involving the Faces task in

Chapter 2. Therefore, in turn, one could argue that I could use the results from Chapter 2 to determine the estimated effect size for my studies in Chapters 3 and 4. However, as pointed out by Brysbaert (2019), given that pilot testing tends to involve small samples, relying on pilot testing data may not be a valid way of estimating effect size.

Instead, for the vast majority of studies, I aimed to recruit as many participants as possible and to meet the 1600 observations per condition criteria recommended by Brysbaert and Stevens (2018). For example, this criteria was met in Experiment 1 of Chapter 5 where I recruited 29 participants and implemented 60 trials per condition. In Chapter 6, my protocol was to recruit enough participants so that Bayesian statistics provided a conclusive result (i.e. at least 'moderate' evidence) on whether the most critical interaction (language group x conflict score) was significant or not. Many of the studies I conducted involved the recruitment of over 100 participants. In particular, I would like to highlight that my synchrony effect studies (Chapters 3 and 4) involved a considerably larger sample size than in most previous related studies. Nevertheless, I acknowledge that the sample sizes may still not be large enough. Through Bayesian analyses, some of my findings were rendered as 'anecdotal', implying a lack of statistical power, and thus, I was unable to generate clear and definitive conclusions. A prominent example of unclear results arises in the regression analyses in Chapter 6. However, it should be highlighted that recruiting large numbers of participants is logistically challenging even in an online setting. For instance, a frequent issue that I encountered with online testing was attrition, especially in multi-part studies where participants may miss an experimental session. To demonstrate, 38% of participants who signed up for my study as described in Chapter 3 missed an experimental session. Therefore, when accounting for attrition, it is likely that you would need to recruit hundreds, or even thousands, of participants to gain enough statistical power for a multi-part study. Nevertheless, I acknowledge that my studies may have rendered clearer outcomes if statistical power had been enhanced.

7.3.4 The use of absolute cut-offs on response latencies

Throughout my thesis, I used absolute cut-offs to remove potential outliers on response latencies. For example, in Chapter 5, I excluded response latencies faster than 150ms and slower than 1,500ms in Experiment 1. The rationale behind the former exclusion is to exclude fast guesses and ensure that the response latencies represent meaningful responses. Researchers have previously argued that response latencies below 100-200ms cannot be interpreted as valid since time is needed to first perceive the presented stimuli and then, physically produce a response (Whelan, 2008). Arguably, defining a long outlier and thus, response latencies which imply inattentiveness is more difficult to identify than short outliers. Indeed, the absolute cut-offs I used were arbitrary. A problem with using absolute cut-offs to detect outliers is that there is a risk of excluding valid reaction times and in turn, the likelihood of producing a Type-II error is increased. When considering this flaw, one could argue that some of my null findings could be partly attributed to how I excluded the data.

An alternative and common approach to identifying outliers is based on standard deviations (e.g. exclude response latencies which are 2.5 standard deviations above or below the mean). A major advantage of this approach over the absolute cut-off approach is that the chances of eliminating genuine response latencies are reduced. However, this approach to detecting outliers is flawed in the sense that the mean and standard deviation are influenced by the outliers. To demonstrate, using this approach could potentially increase the chances of Type-I errors if the outliers are not evenly spread across the conditions (Cousineau & Chartier, 2010). For instance, there may be more long outliers in condition X than in condition Y and therefore, the chances of observing a significant difference in the means of the two conditions are elevated. Additionally, this approach may fail to detect outliers by having both short and long outliers or extreme outliers present in the data which increases the standard deviation. A perhaps crude illustration of this flaw of the standard deviation approach is imagining if I possessed observation values: 1, 2, 20, 25, 26, 28 and 100. The mean would be 28.9 and the standard deviation would be 30.8. By using the 2.5 standard deviation criterion, 100 would not be identified as an outlier. This issue becomes especially prevalent when the sample size is small and thus, outliers could potentially be highly influential.

7.4 Future research

Considering that I was able to capture the S-S and S-R effects in mouse trajectories (see Chapter 5), one could use the mouse tracking paradigm in studies investigating the bilingual advantage. If we assume that more sensitive measures are required to identify the subtle performance differences between monolinguals and bilinguals, then perhaps the bilingual advantage may emerge more clearly using mouse tracking than in conventional keypress measures. To my knowledge, only a handful of studies have investigated the bilingual advantage using mouse tracking. One of the first studies to use mouse tracking was Incera and McLennan (2016) who compared bilinguals' performance on the Stroop task with monolinguals. Ultimately, they found equivocal average response times between the two groups. However, when focussing on data elicited by mouse tracking, their findings suggested that bilinguals initiated their movements later than monolinguals did. Further, when comparing mouse trajectories, bilinguals made fewer deflections towards the incorrect response than monolinguals. This implies that after response initiation, bilinguals moved their mouse faster towards the correct response than monolinguals. The authors suggested that this result indicates that, although they had slower initiation times, bilinguals made more efficient responses than monolinguals. As a consequence, the authors concluded that bilinguals were 'experts' at managing information and thus, coined the term 'expertise signature'. The 'expertise signature' has recently been reproduced by Damian et al. (2019). They found that Chinese-English bilinguals initiated their mouse movements slower than English monolinguals in cognitive tasks including the Stroop and Flanker. Considering these findings, a potential next step to Chapter 6 is to replicate it but use a mouse tracking paradigm and determine whether the expertise signature appears more prominently for S-S trials than S-R ones as Blumenfeld and Marian (2014) would predict, or more clearly for tasks involving 'interference control' than 'response inhibition' as claimed by Luk et al. (2010).

Given that keypress experiments generally seem to translate well to online testing, an interesting area to investigate is whether this is also the case for mouse tracking studies. The software package Gorilla was used throughout most of the studies reported in this thesis, and it does indeed implement a feature where participants' mouse trajectories can be tracked. However, in Chapter 5, I decided to conduct the mouse tracking paradigm in-person since the implications of adapting this paradigm to an online setting are understudied. I would expect that implementing mouse tracking in an online experiment poses more challenges than key press studies do. As mentioned, there is limited control over hardware when conducting online study and this may be particularly problematic when using mouse tracking. Prior research suggests that the way in which a mouse tracking study is set up (e.g., with regard to cursor speed) substantially influences the results observed (e.g., Grage et al., 2019; Kieslich et al., 2020). In Gorilla, it is possible to restrict the devices used to run the study, and this is useful for deterring participants from running the study on a tablet, for example. However, there are no restrictions on the devices used to move the cursor and thus, participants would be able to complete the study using a touchpad rather than a computer mouse. Differences in result may emerge between using a touchpad and a computer mouse to complete a mouse tracking study (Wirth, Foerster, Kunde & Pfister, 2020). For instance, the former may evoke straighter responses than the latter. This is because less physical movement is required to move a cursor from the bottom to the top of the screen via a touchpad than a computer. Future research could explore the similarities and differences in results between online and in-person mouse tracking and in turn, identify the suitability of using mouse tracking in an online setting.

7.5 Concluding remarks

To conclude, this thesis aimed to investigate intra- and inter-individual factors which may influence cognitive inhibition. To explore these factors, I mainly used the Faces task which provides a measure of various subcomponents of this construct. Chapters 3 and 4 implied that cognitive inhibition of either young or older adults is not subject to potential synchrony effects (interaction of time of testing and individual chronotype). As a consequence, I was able to ignore synchrony effects when exploring a potential bilingual advantage in cognitive inhibition in Chapter 6. Overall, I failed to find convincing support for the claim that bilingualism positively influences cognitive inhibition, and this was the case for various subcomponents of inhibition. However, given the results of Bayesian statistical analyses, I was unable to draw definitive conclusions regarding the question of whether certain aspects of bilingualism are more influential to inhibition than others. Finally, the results from Chapter 5 suggested that mouse tracking may be a promising tool that offers intriguing insight into how cognitive inhibition operates in various conflict-based situations, and which could better capture the possible beneficial effects of bilingualism.

References

Abrahan, V. D., Shifres, F., & Justel, N. (2019). Cognitive benefits from a musical activity in older adults. *Frontiers in Psychology*, 10, 652. https://doi.org/10.3389/fpsyg.2019.00652

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, 20(3), 242-275. https://doi.org/10.1016/j.jneuroling.2006.10.003

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., Cappa, S. F., & Costa, A. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex*, 22(9), 2076-2086. https://doi.org/10.1093/cercor/bhr287

Anderson, J. A. E., Campbell, K. L., Amer, T., Grady, C. L., & Hasher, L. (2014). Timing is everything: Age differences in the cognitive control network are modulated by time of day. *Psychology and Aging*, 29(3), 648-657. https://doi.org/10.1037/a0037243

Anderson, J. A. E., Mak, L., Chahi, A. K., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, 50(1), 250-263. https://doi.org/10.3758/s13428-017-0867-9

Ansaldo, A. I., Ghazi-Saidi, L., & Adrover-Roig, D. (2015). Interference control in elderly bilinguals: Appearances can be misleading. *Journal of Clinical and Experimental Neuropsychology*, 37(5), 455-470. https://doi.org/10.1080/13803395.2014.990359

Anwyl-Irvine, A. L., Massonnie, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388-407. https://doi.org/10.3758/s13428-019-01237-x

Bailey, S. L., & Heitkemper, M. M. (2001). Circadian rhythmicity of cortisol and body temperature: Morningness-eveningness effects. *Chronobiology International*, 18(2), 249-261. https://doi.org/10.1081/cbi-100103189

Barner, C., Schmid, S. R., & Diekelmann, S. (2019). Time-of-day effects on prospective memory. *Behavioural Brain Research*, 376, Article 112179. https://doi.org/10.1016/j.bbr.2019.112179

Bedard, A.-C., Nichols, S., Barbosa, J. A., Schachar, R., Logan, G. D., & Tannock, R. (2002). The development of selective inhibitory control across the life span. *Developmental Neuropsychology*, 21(1), 93-111. https://doi.org/10.1207/S15326942DN2101_5

Bennett, C. L., Petros, T. V., Johnson, M., & Ferraro, F. R. (2008). Individual differences in the influence of time of day on executive functions. *American Journal of Psychology*, 121(3), 349-361. https://doi.org/10.2307/20445471

Bialystok, E. (2016). How hazy views become full pictures. *Language Cognition and Neuroscience*, 31(3), 328-330. https://doi.org/10.1080/23273798.2015.1074255

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, 143(3), 233-262. https://doi.org/10.1037/bul0000099

Bialystok, E. (2020). Null results in bilingualism research: What they tell us and what they don't. *Journal of Multilingual Theories and Practices*, 1(1), 8-22.

Bialystok, E. (2021). Bilingualism as a slice of Swiss cheese. *Frontiers in Psychology*, 12. https://doi.org/10.3389/fpsyg.2021.769323

Bialystok, E., & DePape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 565-574. https://doi.org/10.1037/a0012735

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Developmental Science*, 7(3), 325-339. https://doi.org/10.1111/j.1467-7687.2004.00351.x

Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition*, 112(3), 494-500. https://doi.org/10.1016/j.cognition.2009.06.014

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, 19(2), 290–303. https://doi.org/10.1037/0882-7974.19.2.290

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, 16(4), 240-250. https://doi.org/10.1016/j.tics.2012.03.001

REFERENCES

Bialystok, E., Craik, F. I. M., & Ryan, J. (2006). Executive control in a modified antisaccade task: Effects of aging and bilingualism. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 32(6), 1341-1354. https://doi.org/10.1037/0278-7393.32.6.1341

Bialystok, E., Craik, F. I. M., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effect of bilingualism on cognitive control in the Simon task: evidence from MEG. *Neuroimage*, 24(1), 40-49. https://doi.org/10.1016/j.neuroimage.2004.09.044

Bialystok, E., Kroll, J. F., Green, D. W., MacWhinney, B., & Craik, F. I. M. (2015). Publication bias and the validity of evidence: What's the Connection? *Psychological Science*, 26(6), 944-946. https://doi.org/10.1177/0956797615573759

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, 9(1), 103-119. https://doi.org/10.1177/13670069050090010701

Bialystok, E., Poarch, G., Luo, L., & Craik, F. I. M. (2014). Effects of bilingualism and aging on executive function and working memory. *Psychology and Aging*, 29(3), 696-705. https://doi.org/10.1037/a0037254

Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in Stimulus-Stimulus inhibition. *Bilingualism: Language and Cognition*, 17(3), 610-629. https://doi.org/10.1017/s1366728913000564

Borella, E., Ludwig, C., Dirk, J., & de Ribaupierre, A. (2011). The influence of time of testing on interference, working memory, processing speed, and vocabulary: Age differences in adulthood. *Experimental Aging Research*, 37(1), 76–107. https://doi.org/10.1080/0361073X.2011.536744

Borges, R., & Lyddy, F. (2023). Language affects endorsement of misconceptions about bilingualism. *International Journal of Bilingual Education and Bilingualism*, 1-10. https://doi.org/10.1080/13670050.2022.2164478

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, 8(12), 539-546. https://doi.org/10.1016/j.tics.2004.10.003

Brysbaert M. (2019). How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. *Journal of Cognition*, 2(1), 16. https://doi.org/10.5334/joc.72

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, 1(1), 9. https://doi.org/10.5334/joc.10

Bugg, J. M., DeLosh, E. L., & Clegg, B. A. (2006). Physical activity moderates time-of-day differences in older adults' working memory performance. *Experimental Aging Research*, 32(4), 431-446. https://doi.org/10.1080/03610730600875833

Cajochen, C., Munch, M., Knoblauch, V., Blatter, K., & Wirz-Justice, A. (2006). Age-related changes in the circadian and homeostatic regulation of human sleep. *Chronobiology International*, 23(1-2), 461-474. https://doi.org/10.1080/07420520500545813

Calderon, C. B., Gevers, W., & Verguts, T. (2018). The unfolding action model of initiation times, movement times, and movement paths. *Psychological Review*, 125(5), 785. https://doi.org/10.1037/rev0000110

Calderon, C.B., Verguts, T., & Gevers, W. (2015). Losing the boundary: Cognition biases action well after action selection. *Journal of Experimental Psychology: General*, 144(4), 737-743. https://doi.org/10.1037/xge0000087

Carpenter, S. M., Chae, R. L., & Yoon, C. (2020). Creativity and aging: Positive consequences of distraction. *Psychology and Aging*, 35, 654-662. https://doi.org/10.1037/pag0000470

Casey, L. S., Chandler, J., Levine, A. S., Proctor, A., & Strolovitch, D. Z. (2017). Intertemporal differences among MTurk workers: Time-based sample variations and implications for online data collection. *SAGE Open*, 7(2), 2158244017712774. https://doi.org/10.1177/2158244017712774

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29(6), 2156-2160. https://doi.org/10.1016/j.chb.2013.05.009

Chandler, J., Paolacci, G., Peer, E., Mueller, P., & Ratliff, K. A. (2015). Using nonnaive participants can reduce effect sizes. *Psychological Science*, 26(7), 1131-1139. https://doi.org/10.1177/0956797615585115

Chaput, J. P., Dutil, C., & Sampasa-Kanyinga, H. (2018). Sleeping hours: what is the ideal number and how does age impact this? *Nature and Science of Sleep*, 10, 421-430. https://doi.org/10.2147/nss.S163071

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1283-1296. https://doi.org/10.1037/0278-7393.26.5.1283

Costa, A., Hernandez, M., Costa-Faidella, J., & Sebastian-Galles, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, 113(2), 135-149. https://doi.org/10.1016/j.cognition.2009.08.001

Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: a review. *International Journal of Psychological Research*, 3(1), 58-67. https://doi.org/10.21500/20112084.844

Cox, S. R., Bak, T. H., Allerhand, M., Redmond, P., Starr, J. M., Deary, I. J., & MacPherson, S. E. (2016). Bilingualism, social cognition and executive functions: A tale of chickens and eggs. *Neuropsychologia*, 91, 299-306. https://doi.org/10.1016/j.neuropsychologia.2016.08.029

Cramer, A. O. J., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R., Waldorp, L. J., & Wagenmakers, E. J. (2016). Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic Bulletin & Review*, 23(2), 640-647. https://doi.org/10.3758/s13423-015-0913-5

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *Plos One*, 8(3), Article e57410. https://doi.org/10.1371/journal.pone.0057410

Cyr, A. A., Romero, K., & Galin-Corini, L. (2021). Web-based cognitive testing of older adults in person versus at home: Within-subjects comparison study. *Jmir Aging*, 4(1), Article e23384. https://doi.org/10.2196/23384

Damian, M. F., Ye, W., Oh, M., & Yang, S. (2019). Bilinguals as "experts"? Comparing performance of mono- to bilingual individuals via a mousetracking paradigm. *Bilingualism: Language and Cognition,* 22(5), 1176-1193. https://doi.org/10.1017/S1366728918000901

Darcy, N. T. (1953). A Review of the Literature on the effects of bilingualism upon the measurement of intelligence. *Journal of Genetic Psychology,* 82(1), 21-57. https://doi.org/10.1080/08856559.1953.10533654

de Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences,* 9(3), 33. https://doi.org/10.3390/bs9030033

de Bruin, A., Dick, A. S., & Carreiras, M. (2021). Clear theories are needed to interpret differences: Perspectives on the bilingual advantage debate. *Neurobiology of Language*, 2(4), 433-451. https://doi.org/10.1162/nol_a_00038

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, 26(1), 99-107. https://doi.org/10.1177/0956797614557866

Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., Penke, L., Rafnsson, S. B., & Starr, J. M. (2009). Age-associated cognitive decline. *British Medical Bulletin,* 92(1), 135-152. https://doi.org/10.1093/bmb/ldp033

Deary, I. J., Liewald, D., & Nissan, J. (2011). A free, easy-to-use, computer-based simple and four-choice reaction time programme: The Deary-Liewald reaction time task. *Behavior Research Methods,* 43(1), 258-268. https://doi.org/10.3758/s13428-010-0024-1

Della Monica, C., Johnsen, S., Atzori, G., Groeger, J. A., & Dijk, D. J. (2018). Rapid eye movement sleep, sleep continuity and slow wave sleep as predictors of cognition, mood, and subjective sleep quality in healthy men and women, Aged 20-84 Years. *Frontiers in Psychiatry*, 9, 255. https://doi.org/10.3389/fpsyt.2018.00255

Di Milia, L., Adan, A., Natale, V., & Randler, C. (2013). Reviewing the psychometric properties of contemporary circadian typology measures. *Chronobiology International*, 30(10), 1261-1271. https://doi.org/10.3109/07420528.2013.817415

REFERENCES

Dumas, J. A. (2015). What Is Normal Cognitive Aging? Evidence from task-based functional neuroimaging. *Current Behavioral Neuroscience Reports*, 2(4), 256-261. https://doi.org/10.1007/s40473-015-0058-x

Egner, T. (2008). Multiple conflict-driven control mechanisms in the human brain. *Trends in Cognitive Sciences*, 12(10), 374-380. https://doi.org/10.1016/j.tics.2008.07.001

Ekstrom, R. B., French, J. W., Harman, M. H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. Princeton, NJ: Educational Testing Service

Erb, C. D., & Marcovitch, S. (2018). Deconstructing the Gratton effect: Targeting dissociable trial sequence effects in children, pre-adolescents, and adults. *Cognition*, 179, 150-162. https://doi.org/10.1016/j.cognition.2018.06.007

Erb, C. D., Moher, J., Sobel, D. M., & Song, J. H. (2016). Reach tracking reveals dissociable processes underlying cognitive control. *Cognition*, 152, 114–126. https://doi.org/10.1016/j.cognition.2016.03.015

Erb, C. D., Smith, K. A., & Moher, J. (2021). Tracking continuities in the flanker task: From continuous flow to movement trajectories. *Attention, Perception, & Psychophysics*, 83(2), 731-747. https://doi.org/10.3758/s13414-020-02154-4

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16, 143-149. https://doi.org/10.3758/BF03203267

Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics*, 25(4), 249-263. https://doi.org/10.3758/BF03198804

Facer-Childs, E. R., Boiling, S., & Balanos, G. M. (2018). The effects of time of day and chronotype on cognitive and physical performance in healthy volunteers. *Sports Medicine - Open*, 4(1), 47. https://doi.org/10.1186/s40798-018-0162-z

Favieri, F., Forte, G., & Casagrande, M. (2019). The Executive Functions in Overweight and Obesity: A Systematic Review of Neuropsychological Cross-Sectional and Longitudinal Studies. *Frontiers in Psychology*, 10. https://doi.org/10.3389/fpsyg.2019.02126

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42(1), 226-241. https://doi.org/10.3758/BRM.42.1.226

Freeman, J., Dale, R., & Farmer, T. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, 2, 59. https://doi.org/10.3389/fpsyg.2011.00059

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, 133(1), 101-135. https://doi.org/10.1037/0096-3445.133.1.101

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, 5(3), 490-495. https://doi.org/10.3758/BF03208827

Fuller-Thomson, E., & Kuh, D. (2014). The healthy migrant effect may confound the link between bilingualism and delayed onset of Alzheimer's disease. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 52, 128-130. https://doi.org/10.1016/j.cortex.2013.08.009

Funes, M. J., Lupiáñez, J., & Humphreys, G. (2010). Sustained vs. transient cognitive control: Evidence of a behavioral dissociation. *Cognition*, 114(3), 338-347. https://doi.org/10.1016/j.cognition.2009.10.007

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 23(27), 9240–9245. https://doi.org/10.1523/JNEUROSCI.23-27-09240.2003

Gathercole, V. C., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Viñas Guasch, N., Roberts, E. J., Hughes, E. K., & Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Front Psychol*, 5, 11. https://doi.org/10.3389/fpsyg.2014.00011

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, 19(5), 847-857. https://doi.org/10.3758/s13423-012-0296-9

Goldrick, M., Putnam, M., & Schwarz, L. (2016). Coactivation in bilingual grammars: A computational account of code mixing. *Bilingualism: Language and Cognition*, 19(5), 857-876. https://doi.org/10.1017/S1366728915000802

Grage, T., Schoemann, M., Kieslich, P. J., & Scherbaum, S. (2019). Lost to translation: How design factors of the mouse-tracking procedure impact the inference from action to cognition. *Attention, Perception, & Psychophysics*, 81(7), 2538-2557. https://doi.org/10.3758/s13414-019-01889-z

Granic, I., Lobel, A., & Engels, R. (2014). The benefits of playing video games. *American Psychologist*, 69(1), 66-78. https://doi.org/10.1037/a0034857

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology (Hove, England)*, 25(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Grosjean, F., & Li, P. (2013). The psycholinguistics of bilingualism. *John Wiley & Sons*.

Guerrieri, R., Nederkoorn, C., & Jansen, A. (2012). Disinhibition is easier learned than inhibition. The effects of (dis)inhibition training on food intake. *Appetite*, 59(1), 96-99. https://doi.org/10.1016/j.appet.2012.04.006

Hahn, C., Cowell, J. M., Wiprzycka, U. J., Goldstein, D., Ralph, M., Hasher, L., & Zelazo, P. D. (2012). Circadian rhythms in executive function during the transition to adolescence: the effect of synchrony between chronotype and time of day. *Developmental Science*, 15(3), 408-416. https://doi.org/10.1111/j.1467-7687.2012.01137.x

Harrell, F. E. (2015). Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Heidelberg New York Dordrecht London: *Springer*.

Hartanto, A., & Yang, H. (2016). Disparate bilingual experiences modulate task-switching advantages: A diffusion-model analysis of the effects of interactional context on switch costs. *Cognition*, 150, 10-19. https://doi.org/10.1016/j.cognition.2016.01.016

Hartanto, A., & Yang, H. (2020). The role of bilingual interactional contexts in predicting interindividual variability in executive functions: A latent variable analysis. *Journal of Experimental Psychology: General*, 149(4), 609–633. https://doi.org/10.1037/xge0000672

Hartigan, J. A., & Hartigan, P. M. (1985). The Dip Test of Unimodality. *The Annals of Statistics*, 13(1), 70–84. https://doi.org/10.1214/aos/1176346577

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning and Motivation*, 22, 193-225. https://doi.org/10.1016/S0079-7421(08)60041-9

Hasher, L., Chung, C., May, C. P., & Foong, N. (2002). Age, time of testing, and proactive interference. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale*, 56(3), 200–207. https://doi.org/10.1037/h0087397

Hasher, L., Lustig, C., & Zacks, R. (2007). Inhibitory mechanisms and the control of attention. In *Variation in Working Memory*. (pp. 227-249). Oxford University Press.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavioural Research Methods*, 50(3), 1166-1186. https://doi.org/10.3758/s13428-017-0935-1

Hernández, M., Costa, A., Fuentes, L. J., Vivas, A. B., & Sebastián-Gallés, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Bilingualism: Language and Cognition*, 13, 315-325. https://doi.org/10.1017/S1366728909990010

Hill, T. D., Angel, J. L., & Balistreri, K. S. (2012). Does the "healthy immigrant effect" extend to cognitive aging? In J. L. Angel, F. Torres-Gil, & K. Markides (Eds.*), Aging, Health, and Longevity in the Mexican-Origin Population* (pp. 19-33). Springer US. https://doi.org/10.1007/978-1-4614-1867-2_3

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: a new approach. *Psychophysiology*, 10(4), 431-436. https://doi.org/10.1111/j.1469-8986.1973.tb00801.x

Hommel, B. (1997). Interactions between stimulus-stimulus congruence and stimulus-response compatibility. *Psychological Research*, 59(4), 248-260. https://doi.org/10.1007/BF00439302

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica*, 136, 189-202. https://doi.org/10.1016/j.actpsy.2010.04.011

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*.

Horne, J. A., & Östberg, O. (1977). Individual differences in human circadian rhythms. *Biological Psychology*, 5(3), 179-190.

Hufner, K., Binetti, C., Hamilton, D. A., Stephan, T., Flanagin, V. L., Linn, J., Labudda, K., Markowitsch, H., Glasauer, S., Jahn, K., Strupp, M., & Brandt, T. (2011). Structural and functional plasticity of the hippocampal formation in professional dancers and slackliners. *Hippocampus*, 21(8), 855-865. https://doi.org/10.1002/hipo.20801

Hugon, J., Msika, E.-F., Queneau, M., Farid, K., & Paquet, C. (2022). Long COVID: cognitive complaints (brain fog) and dysfunction of the cingulate cortex. *Journal of Neurology*, 269(1), 44-46. https://doi.org/10.1007/s00415-021-10655-x

Incera, S., & McLennan, C. T. (2016). Mouse tracking reveals that bilinguals behave like experts. *Bilingualism: Language and Cognition*, 19(3), 610-620. https://doi.org/10.1017/S1366728915000218

Intons-Peterson, M. J., Rocchi, P., West, T., McLellan, K., & Hackney, A. (1999). Age, testing at preferred or nonpreferred times (testing optimality), and false memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 25(1), 23-40. https://doi.org/10.1037/0278-7393.25.1.23

Intons-Peterson, M. J., Rocchi, P., West, T., McLellan, K., & Hackney, A. (1998). Aging, optimal testing times, and negative priming. *Journal of Experimental Psychology: Learning Memory and Cognition*, 24(2), 362-376. https://doi.org/10.1037/0278-7393.24.2.362

Ioannidis, K., Hook, R., Wickham, K., Grant, J. E., & Chamberlain, S. R. (2019). Impulsivity in gambling disorder and problem gambling: a meta-analysis. *Neuropsychopharmacology*, 44(8), 1354-1361. https://doi.org/10.1038/s41386-019-0393-9

Jylkkä, J., Soveri, A., Laine, M., & Lehtonen, M. (2020). Assessing bilingual language switching behavior with Ecological Momentary Assessment. *Bilingualism: Language and Cognition*, 23(2), 309-322. doi:10.1017/S1366728918001190

Kałamała, P., Szewczyk, J., Chuderski, A., Senderecka, M., & Wodniecka, Z. (2020). Patterns of bilingual language use and response inhibition: A test of the adaptive control hypothesis. *Cognition*, 204, 104373. https://doi.org/10.1016/j.cognition.2020.104373

Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J. M., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: A practical guide to implementation and analysis 1. In *A Handbook of Process Tracing Methods* (pp. 111-130). Routledge.

Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2020). Design factors in mouse-tracking: What makes a difference? *Behavior Research Methods*, 52(1), 317-341. https://doi.org/10.3758/s13428-019-01228-y

Kim, S., Hasher, L., & Zacks, R. T. (2007). Aging and a benefit of distractibility. *Psychonomic Bulletin & Review*, 14(2), 301-305. https://doi.org/10.3758/BF03194068

Knight, M., & Mather, M. (2013). Look out—It's your off-peak time of day! Time of day matters more for alerting than for orienting or executive attention. *Experimental Aging Research*, 39(3), 305-321. https://doi.org/10.1080/0361073X.2013.779197

Kornblum, S. (1994). The way irrelevant dimensions are processed depends on what they overlap with: The case of Stroop-and Simon-like stimuli. *Psychological Research*, 56(3), 130-135. https://doi.org/10.1007/BF00419699

Kornblum, S., Stevens, G. T., Whipple, A., & Requin, J. (1999). The effects of irrelevant stimuli: 1. The time course of stimulus–stimulus and stimulus–response consistency effects with Stroop-like stimuli, Simon-like tasks, and their factorial combinations. *Journal of Experimental Psychology: Human Perception and Performance*, 25(3), 688. https://doi.org/10.1037/0096-1523.25.3.688

Kousaie, S., & Phillips, N. A. (2012). Ageing and bilingualism: Absence of a “bilingual advantage” in Stroop interference in a non-immigrant sample. *Quarterly Journal of Experimental Psychology*, 65(2), 356-369. https://doi.org/10.1080/17470218.2011.604788

REFERENCES

Kousaie, S., Laliberté, C., López Zunini, R., & Taler, V. (2015). A behavioral and electrophysiological investigation of the effect of bilingualism on lexical ambiguity resolution in young adults. *Frontiers in Human Neuroscience, 9*, 682. https://doi.org/10.3389/fnhum.2015.00682

Kunde, W., & Stöcker, C. (2002). A Simon effect for stimulus-response duration. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 55A(2), 581-592. https://doi.org/10.1080/02724980143000433

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning Memory and Cognition*, 37(4), 952-965. https://doi.org/10.1037/a0023217

Lai, G., & O'Brien, B. A. (2020). Examining Language Switching and Cognitive Control Through the Adaptive Control Hypothesis. *Frontiers in Psychology*, 11, 1171. https://doi.org/10.3389/fpsyg.2020.01171

Lara, T., Madrid, J. A., & Correa, A. (2014). The vigilance decrement in executive function is attenuated when individual chronotypes perform at their optimal time of day. *Plos One*, 9(2), Article e88820. https://doi.org/10.1371/journal.pone.0088820

Lehmann, C. A., Marks, A. D. G., & Hanstock, T. L. (2013). Age and synchrony effects in performance on the Rey Auditory Verbal Learning Test. *International Psychogeriatrics*, 25(4), 657-665. https://doi.org/10.1017/s1041610212002013

Lemhofer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44(2), 325-343. https://doi.org/10.3758/s13428-011-0146-0

Levandovski, R., Sasso, E., & Hidalgo, M. P. (2013). Chronotype: a review of the advances, limits and applicability of the main instruments used in the literature to assess human phenotype. *Trends in Psychiatry and Psychotherapy*, 35, 3-11. https://doi.org/10.1590/s2237-60892013000100002

Li, K. Z. H., Hasher, L., Jonas, D., Rahhal, T. A., & May, C. P. (1998). Distractibility, circadian arousal, and aging: A boundary condition? *Psychology and Aging*, 13, 574-583. https://doi.org/10.1037/0882-7974.13.4.574

Li, Q., Nan, W., Wang, K., & Liu, X. (2014). Independent processing of stimulus-stimulus and stimulus-response conflicts. *PLoS One*, 9(2), e89249. https://doi.org/10.1371/journal.pone.0089249

Luk, G., Anderson, J. A. E., Craik, F. I. M., Grady, C., & Bialystok, E. (2010). Distinct neural correlates for two types of inhibition in bilinguals: Response inhibition versus interference suppression. *Brain and Cognition*, 74(3), 347-357. https://doi.org/10.1016/j.bandc.2010.09.004

Luk, G., Bialystok, E., Craik, F. I., & Grady, C. L. (2011). Lifelong bilingualism maintains white matter integrity in older adults. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 31(46), 16808–16813. https://doi.org/10.1523/JNEUROSCI.4563-11.2011

Luk, G., De Sa, E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, 14(4), 588-595. https://doi.org/10.1017/s1366728911000010

Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a "new view". In *Inhibition in Cognition*. (pp. 145-162). American Psychological Association. https://doi.org/10.1037/11587-008

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

Macrae, C. N., Hood, B. M., Milne, A. B., Rowe, A. C., & Mason, M. F. (2002). Are you looking at me? Eye gaze and person perception. *Psychological Science*, 13(5), 460-464. https://doi.org/10.1111/1467-9280.00481

Magnuson, J. S. (2005). Moving hand reveals dynamics of thought. *Proceedings of the National Academy of Sciences*, 102(29), 9995-9996. https://doi.org/10.1073/pnas.0504413102

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America*, 97(8), 4398-4403. https://doi.org/10.1073/pnas.070039597

Marinkovic, K., Rickenbacher, E., & Azma, S. (2012). Effects of Alcohol Intoxication on Response Conflict in a Flanker Task. *Journal of Addiction Research & Therapy*, Suppl 3, 002. https://doi.org/10.4172/2155-6105.S3-002

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 11(1), 81-93. https://doi.org/10.1017/S1366728907003227

May, C. P. (1999). Synchrony effects in cognition: The costs and a benefit. *Psychonomic Bulletin & Review*, 6(1), 142-147. https://doi.org/10.3758/bf03210822

May, C. P., & Hasher, L. (1998). Synchrony effects in inhibitory control over thought and action. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2), 363-379. https://doi.org/10.1037/0096-1523.24.2.363

May, C. P., & Hasher, L. (2017). Synchrony affects performance for older but not younger neutral-type adults. *Timing & Time Perception*, 5(2), 129-148. https://doi.org/10.1163/22134468-00002087

May, C. P., Hasher, L., & Healey, K. (2023). For Whom (and When) the Time Bell Tolls: Chronotypes and the Synchrony Effect. *Perspectives on Psychological Science*, 18(6), 1520-1536. https://doi.org/10.1177/17456916231178553

May, C. P., Hasher, L., & Stoltzfus, E. R. (1993). Optimal time of day and the magnitude of age differences in memory. *Psychological Science*, 4(5), 326-330. https://doi.org/10.1111/j.1467-9280.1993.tb00573.x

May, C. P., Zacks, R. T., Hasher, L., & Multhaup, K. S. (1999). Inhibition in the processing of garden-path sentences. *Psychology and Aging*, 14(2), 304-313. https://doi.org/10.1037//0882-7974.14.2.304

Maylor, E. A., & Badham, S. P. (2018). Effects of time of day on age-related associative deficits. *Psychology and Aging*, 33(1), 7-16. https://doi.org/10.1037/pag0000199

McClelland, J. L. (1979). On the time relations of mental processes: an examination of systems of processes in cascade. *Psychological Review*, 86(4), 287. https://doi.org/10.1037/0033-295X.86.4.287

Meier, M. E., & Kane, M. J. (2015). Carving executive control at its joints: Working memory capacity predicts stimulus-stimulus, but not stimulus-response, conflict. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(6), 1849–1872. https://doi.org/10.1037/xlm0000147

Mittelstädt, V., Miller, J., Leuthold, H., Mackenzie, I. G., & Ulrich, R. (2022). The time-course of distractor-based activation modulates effects of speed-accuracy tradeoffs in conflict tasks. *Psychonomic Bulletin & Review*, 29(3), 837-854. https://doi.org/10.3758/s13423-021-02003-x

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason*, 4(2), 61-64. https://doi.org/10.20982/tqmp.04.2.p061

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, 10(6), 719-726. https://doi.org/10.1111/j.1467-7687.2007.00623.x

Murphy, K. J., West, R., Armilio, M. L., Craik, F. I. M., & Stuss, D. T. (2007). Word-list-learning performance in younger and older adults: Intra-individual performance variability and false memory. *Aging Neuropsychology and Cognition*, 14(1), 70-94. https://doi.org/10.1080/138255890969726

Natale, V., & Alzani, A. (2001). Additional validity evidence for the composite scale of morningness. *Personality and Individual Differences*, 30(2), 293-301. https://doi.org/10.1016/S0191-8869(00)00046-5

Nebel, L. E., Howell, R. H., Krantz, D. S., Falconer, J. J., Gottdiener, J. S., & Gabbay, F. H. (1996). The circadian variation of cardiovascular stress levels and reactivity: Relationship to individual differences in morningness eveningness. *Psychophysiology*, 33(3), 273-281. https://doi.org/10.1111/j.1469-8986.1996.tb00424.x

Nichols, E. S., Gao, Y., Fregni, S., Liu, L., & Joanisse, M. F. (2021). Individual differences in representational similarity of first and second languages in the bilingual brain. *Human brain mapping*, 42(16), 5433-5445. https://doi.org/10.1002/hbm.25633

Ngo, K. W. J., & Hasher, L. (2017). Optimal testing time for suppression of competitors during interference resolution. *Memory*, 25(10), 1396-1401. https://doi.org/10.1080/09658211.2017.1309437

Nichols, E. S., Gao, Y., Fregni, S., Liu, L., & Joanisse, M. F. (2021). Individual differences in representational similarity of first and second languages in the bilingual brain. *Human Brain Mapping*, 42(16), 5433-5445. https://doi.org/10.1002/hbm.25633

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental Science*, 8(1), 74-87. https://doi.org/10.1111/j.1467-7687.2005.00394.x

Paap, K. (2019). The Bilingual Advantage Debate. In *The Handbook of the Neuroscience of Multilingualism* (pp. 701-735). https://doi.org/10.1002/9781119387725.ch34

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66(2), 232-258. https://doi.org/10.1016/j.cogpsych.2012.12.002

Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology*, 5. https://doi.org/10.3389/fpsyg.2014.00962

Paap, K. R., Anders-Jefferson, R., Mikulinsky, R., Masuda, S., & Mason, L. (2019). On the encapsulation of bilingual language control. *Journal of Memory and Language*, 105, 76-92. https://doi.org/10.1016/j.jml.2018.12.001

Paap, K. R., Anders-Jefferson, R., Zimiga, B., Mason, L., & Mikulinsky, R. (2020). Interference scores have inadequate concurrent and convergent validity: Should we stop using the flanker, Simon, and spatial Stroop tasks? *Cognitive Research: Principles and Implications*, 5(1), 1-27. https://doi.org/10.1186/s41235-020-0207-y

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, 69, 265-278. https://doi.org/10.1016/j.cortex.2015.04.014

Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R., & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology*, 29(2), 89-112. https://doi.org/10.1080/20445911.2016.1248436

Panjeh, S., Pompeia, S., Archer, S. N., Pedrazzoli, M., von Schantz, M., & Cogo-Moreira, H. (2021). What are we measuring with the morningness–eveningness questionnaire? Exploratory factor analysis across four samples from two countries. *Chronobiology International*, 38(2), 234-247. https://doi.org/10.1080/07420528.2020.1815758

Pashler, H., Carrier, M., & Hoffman, J. (1993). Saccadic eye movements and dual-task interference. *The Quarterly Journal of Experimental Psychology Section A*, 46(1), 51-82. https://doi.org/10.1080/14640749308401067

Poarch, G. J., & Krott, A. (2019). A bilingual advantage? An appeal for a change in perspective and recommendations for future research. *Behavioral Sciences*, 9(9), 95. https://doi.org/10.3390/bs9090095

Pornpitakpan, C. (1998). Psychometric properties of the composite scale of morningness: a shortened version. *Personality and Individual Differences*, 25(4), 699-709. https://doi.org/10.1016/S0191-8869(98)80002-0

Pratte, M. S. (2021). Eriksen flanker delta plot shapes depend on the stimulus. *Attention, Perception, & Psychophysics*, 83(2), 685-699. https://doi.org/10.3758/s13414-020-02166-0

Pratte, M. S., Rouder, J. N., Morey, R. D., & Feng, C. (2010). Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception, & Psychophysics*, 72(7), 2013-2025. https://doi.org/10.3758/app.72.7.2013

Prickett, C., Brennan, L., & Stolwyk, R. (2015). Examining the relationship between obesity and cognitive function: A systematic literature review. *Obesity Research & Clinical Practice*, 9(2), 93-113. https://doi.org/10.1016/j.orcp.2014.05.001

Prior, A., & Macwhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, 13(2), 253-262. https://doi.org/10.1017/S1366728909990526

R Core Team (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/.

Rabi, R., Chow, R., Paracha, S., Hasher, L., Gardner, S., Anderson, N. D., & Alain, C. (2022a). The effects of aging and time of day on inhibitory control: An event-related potential study. *Frontiers in Aging Neuroscience*, 14, 821043. https://doi.org/10.3389/fnagi.2022.821043

Rabi, R., Chow, R., Paracha, S., Hasher, L., Gardner, S., Anderson, N. D., & Alain, C. (2022b). Time of day effects on inhibitory functioning: Cognitive and neural evidence of sundowning in amnestic mild cognitive impairment. *Journal of Alzheimer's Disease*, 90, 869-890. https://doi.org/10.3233/JAD-220580

Randler, C., & Engelke, J. (2019). Gender differences in chronotype diminish with age: a meta-analysis based on morningness/chronotype questionnaires. *Chronobiology International*, 36(7), 888-905. https://doi.org/10.1080/07420528.2019.1585867

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, 20(4), 260-281. https://doi.org/10.1016/j.tics.2016.01.007

Rauscher, F. H., & Hinton, S. C. (2011). Music instruction and its diverse extra-musical benefits. *Music Perception: An Interdisciplinary Journal*, 29(2), 215–226. https://doi.org/10.1525/mp.2011.29.2.215

Raven, J., Raven, J. C., & Court, J. H. (1977). Manual for Raven's progressive matrices and vocabulary scales: Advanced progressive matrices, sets I & II. *Psychological Press*.

Rey-Mermet, A. (2020). Finding an interaction between Stroop congruency and flanker congruency requires a large congruency effect: A within-trial combination of conflict tasks. *Attention, Perception, & Psychophysics*, 82, 2271–2301. https://doi.org/10.3758/s13414-019-01914-1

Rey-Mermet, A., & Gade, M. (2018). Inhibition in aging: What is preserved? What declines? A meta-analysis. *Psychonomic Bulletin & Review*, 25(5), 1695-1716. https://doi.org/10.3758/s13423-017-1384-7

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning Memory and Cognition*, 44(4), 501-526. https://doi.org/10.1037/xlm0000450

Rodriguez-Fornells, A., Kramer, U. M., Lorenzo-Seve, U., Festman, J., & Munte, T. F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology*, 3, Article 388. https://doi.org/10.3389/fpsyg.2011.00388

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological Review*, 110(1), 88-125. https://doi.org/10.1037/0033-295X.110.1.88

Roenneberg, T., Kuehnle, T., Juda, M., Kantermann, T., Allebrandt, K., Gordijn, M., & Merrow, M. (2007). Epidemiology of the human circadian clock. *Sleep Medicine Reviews*, 11(6), 429-438. https://doi.org/10.1016/j.smrv.2007.07.005

Rothen, N., & Meier, B. (2016). Time of day affects implicit memory for unattended stimuli. *Consciousness and Cognition*, 46, 1-6. https://doi.org/10.1016/j.concog.2016.09.012

Rowe, G., Hasher, L., & Turcotte, J. (2009). Age and synchrony effects in visuospatial working memory. *The Quarterly Journal of Experimental Psychology*, 62(10), 1873-1880. https://doi.org/10.1080/17470210902834852

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, 8, 535-543. https://doi.org/10.1037/0894-4105.8.4.535

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological psychology*, 54(1), 35-54. https://doi.org/10.1016/S0301-0511(00)00052-1

Salthouse, T. A. (2019). Trajectories of normal cognitive aging. *Psychology and Aging*, 34(1), 17-24. https://doi.org/10.1037/pag0000288

Samuel, S., Roehr-Brackin, K., Pak, H., & Kim, H. (2018). Cultural effects rather than a bilingual advantage in cognition: A review and an empirical study. *Cognitive Science*, 42(7), 2313-2341. https://doi.org/10.1111/cogs.12672

Sanchez-Azanza, V. A., López-Penadés, R., Buil-Legaz, L., Aguilar-Mediavilla, E., & Adrover-Roig, D. (2017). Is bilingualism losing its advantage? A bibliometric approach. *Plos One,* 12(4), e0176151. https://doi.org/10.1371/journal.pone.0176151

Sanders, A.F. (1980). Stage analysis of reaction processes. In G.E. Stelmach & J. Requin (Eds), *Tutorials in Motor Behavior*, pp. 331-354. Amsterdam: North-Holland.

Scherbaum, S., & Dshemuchadse, M. (2020). Psychometrics of the continuous mind: Measuring cognitive sub-processes via mouse tracking. *Memory & Cognition*, 48(3), 436-454. https://doi.org/10.3758/s13421-019-00981-x

Scherbaum, S., Dshemuchadse, M., Fischer, R., & Goschke, T. (2010). How decisions evolve: The temporal dynamics of action selection. *Cognition*, 115(3), 407-416. https://doi.org/10.1016/j.cognition.2010.02.004

Schlaug, G. (2015). Musicians and music making as a model for the study of brain plasticity. In E. Altenmuller, S. Finger, & F. Boller (Eds.), *Music, Neurology, and Neuroscience: Evolution, the Musical Brain, Medical Conditions, and Therapies* (Vol. 217, pp. 37-55). https://doi.org/10.1016/bs.pbr.2014.11.020

Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology*, 24(7), 755-789. https://doi.org/10.1080/02643290701754158

Schmidt, C., Peigneux, P., Cajochen, C., & Collette, F. (2012). Adapting test timing to the sleep-wake schedule: Effects on diurnal neurobehavioral performance changes in young evening and older morning chronotypes. *Chronobiology International*, 29(4), 482-490. https://doi.org/10.3109/07420528.2012.658984

Schmidt, J. R., Crump, M. J. C., Cheesman, J., & Besner, D. (2007). Contingency learning without awareness: Evidence for implicit control. *Consciousness and Cognition*, 16(2), 421-435. https://doi.org/10.1016/j.concog.2006.06.010

Schoemann, M., O’Hora, D., Dale, R., & Scherbaum, S. (2020). Using mouse cursor tracking to investigate online cognition: Preserving methodological ingenuity while moving toward reproducible science. *Psychonomic Bulletin & Review*, 1-22. https://doi.org/10.3758/s13423-020-01851-3

Schroeder, S. R., Marian, V., Shook, A., & Bartolotti, J. (2016). Bilingualism and musicianship enhance cognitive control. *Neural Plasticity*, 2016, Article 4058620. https://doi.org/10.1155/2016/4058620

Schwartz, A. I., Kroll, J. F., & Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes*, 22(1), 106-129. https://doi.org/10.1080/01690960500463920

Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods*, 49(4), 1241-1260. https://doi.org/10.3758/s13428-016-0783-4

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, 51, 300-304. https://doi.org/10.1037/h0020586

Singh, N., & Mishra, R. K. (2013). Second language proficiency modulates conflict-monitoring in an oculomotor Stroop task: evidence from Hindi-English bilinguals. *Frontiers in Psychology*, 4, Article 322. https://doi.org/10.3389/fpsyg.2013.00322

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2016). afex: Analysis of factorial experiments. R package version 0.16-1.

Smith, J. L., Mattick, R. P., Jamadar, S. D., & Iredale, J. M. (2014). Deficits in behavioural inhibition in substance abuse and addiction: A meta-analysis. *Drug and Alcohol Dependence*, 145, 1-33. https://doi.org/10.1016/j.drugalcdep.2014.08.009

Song, J. H., & Nakayama, K. (2008). Target selection in visual search as revealed by movement trajectories. *Vision Research*, 48(7), 853–861. https://doi.org/10.1016/j.visres.2007.12.015

Song, J., & Stough, C. (2000). The relationship between morningness-eveningness, time-of-day, speed of information processing, and intelligence. *Personality and Individual Differences*, 29(6), 1179-1190. https://doi.org/10.1016/s0191-8869(00)00002-7

Soveri, A., Rodriguez-Fornells, A., & Laine, M. (2011). Is there a relationship between language switching and executive functions in bilingualism? Introducing a within group analysis approach. *Frontiers in Psychology*, 2, 183. https://doi.org/10.3389/fpsyg.2011.00183

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 461-479. https://doi.org/10.1037/0096-1523.22.2.461

Spivey, M. J., & Dale, R. (2004). On the continuity of mind: toward a dynamical account of cognition. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory*, Vol. 45, pp. 87–142). Elsevier Academic Press.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276-315. https://doi.org/10.1016/0001-6918(69)90055-9

Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences*, 21(10), 736-748. https://doi.org/10.1016/j.tics.2017.06.007

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643-662. https://doi.org/10.1037/h0054651

Taillard, J., Philip, P., Chastang, J.-F., & Bioulac, B. (2004). Validation of Horne and Ostberg Morningness-Eveningness Questionnaire in a middle-aged population of French workers. *Journal of Biological Rhythms*, 19(1), 76-86. https://doi.org/10.1177/0748730403259849

Tao, L., Marzecová, A., Taft, M., Asanowicz, D., & Wodniecka, Z. (2011). The efficiency of attentional networks in early and late bilinguals: The role of age of acquisition. *Frontiers in Psychology*, 2. https://doi.org/10.3389/fpsyg.2011.00123

van Heuven, W. J. B., & Dijkstra, T. (2010). Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain Research Reviews*, 64(1), 104-122. https://doi.org/10.1016/j.brainresrev.2010.03.002

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods*, 49(2), 653-673. https://doi.org/10.3758/s13428-016-0721-5

Verhaeghen, P. (2011). Aging and executive control: Reports of a demise greatly exaggerated. *Current Directions in Psychological Science*, 20, 174-180. https://doi.org/10.1177/0963721411408772

Verhaeghen, P. (2014). The elements of cognitive aging: Meta-analyses of age-related differences in processing speed and their consequences. *Oxford University Press*.

von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology: General*, 145(2), 246-258. https://doi.org/10.1037/xge0000120

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1), 35-57. https://doi.org/10.3758/s13423-017-1343-3

Wang, K., Li, Q., Zheng, Y., Wang, H., & Liu, X. (2014). Temporal and spectral profiles of stimulus–stimulus and stimulus–response conflict processing. *Neuroimage*, 89, 280-288. https://doi.org/10.1016/j.neuroimage.2013.11.045

Ware, A. T., Kirkovski, M., & Lum, J. A. G. (2020). Meta-analysis reveals a bilingual advantage that is dependent on task and age. *Frontiers in Psychology*, 11, Article 1458. https://doi.org/10.3389/fpsyg.2020.01458

Weaver Cargin, J., Collie, A., Masters, C., & Maruff, P. (2008). The nature of cognitive complaints in healthy older adults with and without objective memory decline. *Journal of Clinical and Experimental Neuropsychology*, 30(2), 245-257. https://doi.org/10.1080/13803390701377829

West, R., Murphy, K. J., Armilio, M. L., Craik, F. I. M., & Stuss, D. T. (2002). Effects of time of day on age differences in working memory. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences*, 57(1), P3-P10. https://doi.org/10.1093/geronb/57.1.P3

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, 58(3), 475–482. https://doi.org/10.1007/BF03395630

Wilckens, K. A., Woo, S. G., Kirk, A. R., Erickson, K. I., & Wheeler, M. E. (2014). Role of sleep continuity and total sleep time in executive function across the adult lifespan. *Psychology and Aging*, 29(3), 658–665. https://doi.org/10.1037/a0037234

Wilson, R. S., Beckett, L. A., Barnes, L. L., Schneider, J. A., Bach, J., Evans, D. A., & Bennett, D. A. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging*, 17(2), 179-193. https://doi.org/10.1037/0882-7974.17.2.179

Wirth, R., Foerster, A., Kunde, W., & Pfister, R. (2020). Design choices: Empirical recommendations for designing two-dimensional finger-tracking experiments. *Behavior Research Methods*, 52(6), 2394-2416. https://doi.org/10.3758/s13428-020-01409-0

Wispinski, N. J., Gallivan, J. P., & Chapman, C. S. (2020). Models, movements, and minds: bridging the gap between decision making and action. *Annals of the New York Academy of Sciences*, 1464(1), 30–51. https://doi.org/10.1111/nyas.13973

Woumans, E., Ceuleers, E., Van der Linden, L., Szmalec, A., & Duyck, W. (2015). Verbal and nonverbal cognitive control in bilinguals and interpreters. *Journal of Experimental Psychology-Learning Memory and Cognition*, 41(5), 1579-1586. https://doi.org/10.1037/xlm0000107

Wulff, D. U., Haslbeck, J. M., Kieslich, P. J., Henninger, F., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: Detecting types in movement trajectories. In D. U. Wulff et al. (Eds)., *A handbook of process tracing methods* (pp. 131-145). Routledge.

Xia, L., Bak, T. H., Sorace, A., & Vega-Mendoza, M. (2022). Interference suppression in bilingualism: Stimulus-Stimulus vs. Stimulus-Response conflict. *Bilingualism: Language and Cognition*, 25(2), 256-268. https://doi.org/10.1017/S1366728921000304

Yang, J., Ru, T., Chen, Q., Mao, T., Ji, Y., & Zhou, G. (2019). The effects of ambient light on task switching depend on the chronotype. *Lighting Research & Technology*, 51(4), 544-556. https://doi.org/10.1177/1477153518777028

Yang, L., Hasher, L., & Wilson, D. E. (2007). Synchrony effects in automatic and controlled retrieval. *Psychonomic Bulletin & Review*, 14(1), 51-56. https://doi.org/10.3758/BF03194027

Yang, L., Hasher, L., & Wilson, D. E. (2007). Synchrony effects in automatic and controlled retrieval. *Psychonomic Bulletin & Review*, 14(1), 51-56. https://doi.org/10.3758/BF03194027

Ye, W., & Damian, M. F. (2023). Effects of conflict in cognitive control: Evidence from mouse tracking. *Quarterly Journal of Experimental Psychology*, 76(1), 54–69. https://doi.org/10.1177/17470218221078265

Yoon, C., May, C. P., & Hasher, L. (1999). Aging, circadian arousal patterns, and cognition. In *Cognition, aging, and self-reports*. (pp. 117-143). Psychology Press/Erlbaum (UK) Taylor & Francis. https://doi.org/10.4324/9780203345115

Zaninotto, P., Batty, G. D., Allerhand, M., & Deary, I. J. (2018). Cognitive function trajectories and their determinants in older people: 8 years of follow-up in the English Longitudinal Study of Ageing. *Journal of Epidemiology and Community Health*, 72(8), 685-694. https://doi.org/10.1136/jech-2017-210116

Appendix A

Appendix A provides a supplementary table for Experiment 2 (Chapter 5).

Table A.1

ANOVAs with factors, congruency (congruent vs incongruent) and conflict type (S-S vs S-R), performed on error rates, initiation times, response latencies and Area Under Curve. Significant Results are bolded.

Errors	df	MSE	F	η^2	p	BF_{10}
Congruency	**1, 27**	**2**	**5.01**	**.156**	**.034**	**1.90**
Conflict type	**1, 27**	**1**	**9.61**	**.263**	**.004**	**0.22**
Congruency x Conflict type	1, 27	5	3.46	.114	.074	13.53
Initiation times	df	MSE	F	η^2	p	
Congruency	**1, 27**	**164**	**5.24**	**.163**	**.030**	**0.624**
Conflict type	**1, 27**	**152**	**43.38**	**.616**	**<.001**	**>100**
Congruency x Conflict type	1, 27	201	1.60	.056	.217	0.09
Response latencies	df	MSE	F	η^2	p	
Congruency	**1, 27**	**940**	**150.23**	**.848**	**<.001**	**>100**
Conflict type	**1, 27**	**724**	**11.99**	**.308**	**.002**	**19.48**
Congruency x Conflict type	**1, 27**	**1151**	**76.62**	**.739**	**<.001**	**>100**
Area under Curve	df	MSE	F	η^2	p	
Congruency	**1, 27**	**4008**	**215.96**	**.889**	**<.001**	**>100**
Conflict type	**1, 27**	**1289**	**73.44**	**.731**	**<.001**	**>100**
Congruency x Conflict type	**1, 27**	**4042**	**162.93**	**.858**	**<.001**	**>100**

Appendix B

Appendix B provides supplementary tables for Experiment 1 and 2 (Chapter 6).

Table B.1

*Faces task. A correlation matrix displaying the relationship between language measures (AoA, LexTALE, Contextual switch, Unintended switch, Switch to L1 and Switch to L2), demographic factors (SES and Raven's score) and task performance (Average RT, Inhibitory Control, Response Suppression and Task Switching). * denotes p < .05 and ** denotes p < .01.*

		SES	Raven's score	AoA	LexTALE	Contextual switch	Unintended switch	Switch to L1	Switch to L2	Average RT	Inhibitory Control	Response Suppression
Raven's score												
	r	.15										
	BF_{10}	0.32										
AoA												
	r	-.13	.09									
	BF_{10}	0.27	0.20									
LexTALE												
	r	.02	-.08	-.12								
	BF_{10}	0.16	0.19	0.24								
Contextual switch												
	r	.21	-.07	-.19	-.18							
	BF_{10}	0.58	0.19	0.47	0.39							
Unintended switch												
	r	.07	.01	-.01	-.10	.54**						
	BF_{10}	0.19	0.16	0.16	0.21	>100						
Switch to L1												
	r	.17	.11	-.15	-.24	.37**	.35**					
	BF_{10}	0.38	0.23	0.32	0.80	9.40	6.38					
Switch to L2												
	r	.40**	.17	-.16	-.17	.51**	.44**	.42**				
	BF_{10}	23.27	0.37	0.35	0.36	>100	66.82	39.13				

		SES	Raven's score	AoA	LexTALE	Contextual switch	Unintended switch	Switch to L1	Switch to L2	Average RT	Inhibitory Control	Response Suppression
Average RT												
	r	-.01	-.07	.00	-.17	.19	.34**	.24	.08			
	BF_{10}	0.16	0.18	0.16	0.36	0.47	5.07	0.81	0.19			
Inhibitory Control												
	r	.02	-.07	-.02	-.10	.00	.04	.11	-.04	.38**		
	BF_{10}	0.16	0.18	0.16	0.21	0.16	0.17	0.23	0.17	12.18		
Response Suppression												
	r	.08	.07	.06	.08	.28**	.26*	.22	.18	.36**	.04	
	BF_{10}	0.20	0.19	0.18	0.19	1.60	1.22	0.69	0.40	8.18	0.17	
Task Switching												
	r	.10	.09	.02	-.23	.16	-.01	.36**	.21	.51**	.28*	.44**
	BF_{10}	0.21	0.21	0.16	0.76	0.32	0.16	7.52	0.55	>100	1.62	69.95

Table B.2
Faces task. ANCOVAs performed on response latencies and errors, separately for each component of cognitive control (inhibitory control, response suppression, task switching) and the factor group (monolinguals vs. bilinguals). Raven's score was entered as a covariate. Significant results are bolded

	Response Latencies				Errors			
Effect	MSE	F	η_p^2	p	MSE	F	η_p^2	p
Inhibitory control								
Group	22,153	0.13	<.001	.716	84	0.11	<.001	.920
Congruency	**1,546**	**43.29**	**.244**	**<.001**	**20**	**22.96**	**.146**	**<.001**
Group × Congruency	1,546	0.39	.003	.668	20	0.01	<.001	.943
Raven	22,153	3.97	.029	.121	**84**	**23.53**	**.149**	**<.001**
Raven x Congruency	1,546	1.13	.008	.484	**20**	**5.37**	**.038**	**.037**
Response suppression								
Group	25,731	0.13	<.001	.954	71	0.48	.004	.564
Response Type	**1,505**	**44.83**	**.251**	**<.001**	**23**	**7.76**	**.055**	**.015**
Group × Response Type	1,505	<0.01	<.001	.954	23	0.33	.002	.564
Raven	25,731	2.71	.020	.256	**71**	**14.15**	**.095**	**.001**
Raven x Response type	1,505	0.02	<.001	.954	23	3.00	.022	.143
Task switching								
Group	23,937	0.15	.001	.868	54	0.11	<.001	.741
Task switch	**879**	**115.15**	**.462**	**<.001**	**13**	**111.40**	**.454**	**<.001**
Group × Task Switch	879	<.001	<.001	.983	13	2.34	.017	.161
Raven	23,937	3.73	.027	.868	**54**	**26.27**	**.164**	**<.001**
Raven x Task Switch	879	0.88	.006	.585	**13**	**14.19**	**.096**	**<.001**

$df_1 = 1$; $df_2 = 134$

Table B.3

*Integrated arrow task. A correlation matrix displaying the relationship between language measures (AoA, LexTALE, Contextual switch, Unintended switch, Switch to L1 and Switch to L2), demographic factors (SES and Raven's score) and task performance (Average RT, overall congruency score, SR conflict score and SS conflict score). * denotes p < .05 and ** denotes p < .01*

		SES	Raven's score	AoA	LexTALE	Contextual switch	Unintended switch	Switch to L1	Switch to L2	Average RT	Overall congruency score	SR conflict score
Raven's score												
	r	.00										
	BF_{10}	0.16										
AoA												
	r	-.17	.15									
	BF_{10}	0.36	0.30									
LexTALE												
	r	.01	-.09	-.10								
	BF_{10}	0.16	0.20	0.21								
Contextual switch												
	r	.22	-.13	-.22	-.18							
	BF_{10}	0.74	0.26	0.67	0.42							
Unintended switch												
	r	-.02	.12	.02	-.15	.50**						
	BF_{10}	0.16	0.24	0.16	0.32	>100						
Switch to L1												
	r	.20	.07	-.16	-.25*	.39**	.28**					
	BF_{10}	0.51	0.18	0.34	1.08	19.7	1.98					
Switch to L2												
	r	.25*	.26*	-.14	-.18	.47**	.52**	.41**				
	BF_{10}	1.06	1.25	0.28	0.44	>100	>100	35.8				
Average RT												
	r	.09	.00	.00	-.23	.20	.22	.24	.05			
	BF_{10}	0.20	0.16	0.16	0.82	0.54	0.65	0.92	0.17			

	SES	Raven's score	AoA	LexTALE	Contextual switch	Unintended switch	Switch to L1	Switch to L2	Average RT	Overall congruency score	SR conflict score
Overall congruency score											
r	-.05	-.21	.08	-.21	.09	.09	.00	-.08	.08		
BF_{10}	0.17	0.64	0.19	0.60	0.20	0.20	0.16	0.19	0.19		
SR conflict score											
r	-.10	-.07	.08	-.21	.03	.07	-.09	-.09	.20	.79**	
BF_{10}	0.22	0.18	0.19	0.58	0.16	0.18	0.20	0.20	0.54	>100	
SS conflict score											
r	.04	-.25*	.04	-.09	.10	.06	.11	-.02	-.13	.64**	.03
BF_{10}	0.17	1.04	0.17	0.20	0.21	0.17	0.23	0.16	0.26	>100	0.16

Table B.4
Integrated arrow task. ANCOVAs performed on response latencies and errors, with congruency (congruent vs incongruent), conflict type (Stimulus-Stimulus vs Stimulus-Response) and language group (monolinguals vs bilinguals). Raven's score was entered as a covariate. Significant results are bolded

	Response Latencies				Errors			
Effect	MSE	F	η_p^2	p	MSE	F	η_p^2	p
Congruency	**302**	**777.82**	**.845**	**<.001**	**23**	**205.92**	**.590**	**<.001**
Conflict Type	**259**	**21.86**	**.133**	**<.001**	**14**	**51.61**	**.265**	**<.001**
Congruency × Conflict Type	**350**	**58.89**	**.292**	**<.001**	**16**	**104.14**	**.421**	**<.001**
Group	14,781	0.14	<.001	.784	61	0.40	.003	.829
Group × Congruency	302	4.23	.039	.056	23	0.26	.002	.845
Group × Conflict Type	259	0.14	.001	.784	14	0.01	<.001	.933
Group × Congruency × Conflict Type	350	0.43	.003	.784	16	0.04	<.001	.933
Raven	14,781	1.60	.011	.381	**61**	**17.95**	**.112**	**<.001**
Raven x Congruency	**302**	**5.87**	**.039**	**.046**	23	2.11	.015	.273
Raven x Conflict Type	259	0.08	<.001	.784	14	0.01	<.001	.933
Raven x Congruency x Conflict Type	350	0.13	<.001	.784	16	3.60	.025	.131

$df_1 = 1$; $df_2 = 143$