UiT The Arctic University of Norway

Department of Psychology, Faculty of Health Sciences

Adjusting the Instruction-Based Congruency Effect to an Explicit cue Requires Practise

Steffen Rygg Aasen

Master's thesis in Psychology – PSY-3900. May 2024

Preface

Approximately three years ago, my journey started. I entered the Student Research Programme in Psychology combined with a master's in psychology. Both have provided invaluable insight into the research process and especially literature on psychological phenomena. The first year was dedicated to the Research Programme where I did research in relation to attention and mind wandering. The last two years have been primarily related to the master thesis, regarding instruction implementation.

For my masters, I wanted to do something other than attention and mind wandering to broaden my horizon. Considering this, I always found Torsten Martiny-Hüenger's work intriguing. I always found the theoretical perspectives he takes on human behaviour to be fascinating. They provide "simple" explanations for why we do what we do. Simple in the sense that they offer concrete explanations for our behaviours. Human behaviours are nevertheless complex, but they are built on the foundation of these "simple" mechanisms. The choice was, therefore, not hard when I was looking for what to do for my masters. I am grateful for the opportunity, conversations, and feedback that Torsten has provided me.

Despite the solo adventure of writing the thesis, the master's programme has not been a lonely road. If I am not mistaken, we have had one of the (if not the) biggest Master classes at UiT so far – in part thanks to the master's turning into an international master. This has provided me, not only new friends, but international friends. The adventure has been filled, not just by solo-writing, but a wide range of social activities and social gatherings. Social gathering not just encompassing the current master (22-24) but also the previous and newest master group. Additionally, the master's offices have offered opportunities to meet other people working at UiT, providing an even broader horizon of friends and colleagues. The collection of people has provided a constant source of socializing at school – during lunch, in

the corridor, in the lab and elsewhere. The social encounters have made the whole journey of the master's bearable and enjoyable.

The unwavering support of my social group has been crucial for my well-being, and I hope my presence has similarly enriched their experiences. I want to express a broad thanks to everyone on the third floor for the social encounters. Thanks to my office friends Samy, Ingebjørg, Ragnhild, Marie, and Ingar for sticking it out these past three years. Additional thanks go out to Otto, Runar, Joakim and Eirik for your social time and energy; Kristian for sticking out with me for well over 4 years, with the absurd sense of humour and experiences we have acquired over these last years. A special thanks to Katrine for listening to my worries and providing comfort. Lastly, a thanks to my family and friends from Bergen for a constant source of support throughout all my years in Tromsø and before. Your support is undoubtedly a big part of this work. Thank you all.

Steffen Rygg Aasen

Torsten Martiny-Hüenger

Adjusting the Instruction-Based Congruency Effect to an Explicit cue Requires Practise

Steffen Rygg Aasen

Supervisor: Torsten Martiny-Huenger

PSY-3900

Master Thesis in Psychology

May 2024

Department of Psychology

Faculty of Health Sciences

UiT – The Arctic University of Norway

Sammendrag

Mennesker viser en evne til raskt å tilpasse atferden til nye instruerte relasjoner. Instruksjoner binder typisk kjente stimuli og responser på nye måter som må bevares for fremtidig implementering, noe som kan føre til utilsiktede effekter på en sekundær oppgave. Denne effekten kan justeres basert på graden av forberedelse for den bevarte instruksjonen. Signaler kan brukes til å eksplisitt signalisere kommende oppgavekrav, slik at graden av forberedelse for de bevarte instruksjonene kan justeres. For å undersøke om forberedelsen kan justeres til et eksplisitt signal, brukte jeg ett design implementert i litteraturen om instruksjonsbasert atferd. Et signal ble presentert et par forsøk før forsøket relatert til den bevarte instruksjonen ble presentert. Jeg antok at effekten på den sekundære oppgaven ville være redusert eller fraværende før signalet, men til stede etter signalet. Analysen viste det forventede mønsteret når man tar hensyn til tid som en interagerende faktor i analysen. Dermed indikerte bare senere deler av eksperimentet det forventede mønsteret, mens de tidligere og midtre delene av eksperimentet ikke gjorde det. Bevisene tyder på at trening er nødvendig for å dempe forberedelsen før signalet, i tråd med tidligere forskning.

Nøkkelord: Instruksjoner, kongruens effekt, oppgave bytting, forberedelse, instruksjons-basert læring

Abstract

Humans display an ability to rapidly adapt behaviour to novelly instructed relations. Instructions typically bind known stimuli and responses in novel ways that must be retained for future implementation, which can induce unintentional effects on a secondary task. This effect can be adjusted based on the degree of preparation for the retained instruction. Cues can be used to explicitly signal upcoming task demands, such that the degree of preparation for the retained instructions can be adjusted. To investigate whether preparation can be adjusted to an explicit cue, I used a novel design implemented in the instruction-based literature. A cue was presented a couple of trials before the trial related to the retained instruction was presented. I hypothesized that the effect on the secondary task would be reduced or absent before the cue but present after the cue. The analysis revealed the expected pattern when considering time as an interacting factor in the analysis. Thus, only the later parts of the experiment indicated the expected pattern, while the earlier and middle parts of the experiment did not. The evidence suggests that practice is necessary to attenuate preparation before the cue, in line with prior research.

Keywords: instructions, congruency effect, task switching, preparation, adaptation, instruction-based research

Adjusting the Instruction-Based Congruency Effect to an Explicit cue Requires Practise

Human beings display an ability to quickly adapt behaviour to new situations. This capacity is often referred to as rapid instruction task learning (RITL; Cole et al., 2017), which is illustrated through the rapid adaptation of behaviour following instructions. In most cases, instructions describe how to respond to a given feature in the environment. The response and feature (stimulus) may be familiar, in the sense that they have been encountered before, but the specific relation between the response and the feature may be novel. For example, in experimental settings, instructions may describe that "if a box appears, then press the j key". The box and the j-key may be familiar to most humans, but the specific relation between the box and the j-key may have never been seen before. Nonetheless, humans can perform the specified relation quickly and successfully without any practice of the specified relation (e.g., Cole et al., 2013; Liefooghe et al., 2012). This quick and successful adaptation suggests that there exists a cognitive mechanism that can adapt behaviour to new situations that significantly enhances behaviour flexibility.

This flexibility, however, can lead to some unfortunate consequences. Research suggests that remembering one novel instructed relation for future implementation can influence the performance on a secondary task (Liefooghe et al., 2012; Liefooghe & De Houwer, 2018; Meiran et al., 2015). While this impact may be trivial in everyday situations, it can be used to uncover cognitive processes underlying cognitive flexibility. Indeed, research suggests that the impact on the secondary task is greater if one is more prepared to implement the novel instructed relation, compared to being less prepared to implement the instructed relation (e.g., Braem et al., 2019; Liefooghe et al., 2013). The task-switching literature indicates that there is a cost of switching between tasks, termed the switch-cost, which can be reduced by providing cues about the upcoming task (Kiesel et al., 2010). However, it remains to be seen whether such cues can be used to adjust the degree of preparation to novel instructed relations because one feature of the adaptive mechanisms to these novel instructed relations is its rigidity (Meiran et al., 2012). In this thesis, therefore, I explore whether explicitly signalling the upcoming task demand can attenuate the degree of preparation before the cue.

Instructions

Instructions are ubiquitous in modern society and rely on our ability to adapt behaviour to novel relations. The instructions typically outline simple steps of procedures to reach a goal. Some goals can be achieved through a single step, while others are more complex and necessitate multiple steps. Often, the instructions outline steps related to familiar situations and behaviours, but the specific relation between them may be novel. Despite the novel relation, humans can quickly and successfully implement the instructed relation (e.g., Liefooghe et al., 2012), bypassing the more laborious trial-and-error learning process (e.g., Ruge et al., 2018). Instructions are therefore instrumental in enabling behavioural flexibility and make it possible for us to engage in a broad range of activities.

To illustrate, consider a person who is making a new dish. Since the dish is new, the person needs to follow a recipe (i.e., instructions) that describe the necessary steps to make the dish. Firstly, the person needs to gather the necessary ingredients and equipment. The ingredients then have to be cut in a certain way and some of the ingredients should be put in a bowl. Thereafter, the ingredients should be stirred together with some liquids. When the mixture turns yellow, the other ingredients should be added and mixed. When the ingredients are well mixed, they can be cooked, and the dish is complete. Fortunately, the person is familiar with all the ingredients, tools, and procedures, but has never stirred these specific ingredients before. Despite the novel relation between the ingredients and the stirring, the person quickly adapts the stirring to the ingredients and successfully creates the new dish.

Components of Instruction Following

The ability to adapt familiar behaviour to unfamiliar situations or novel relations relies on a couple of important components. Firstly, the instructions must be described in a language that is known. For example, in the cooking example above, for the person to successfully follow the instructions, they must be described in a language the person knows (e.g., English) and not in a different language (e.g., Chinese). Otherwise, the instructions could not be interpreted and the attempt to create the dish would highly likely fail. Another crucial component is to remember the elements of the instructions. For example, the person must remember which tools and ingredients to retrieve. Additionally, since the relation between the ingredients and the stirring is novel, the relation must first be established through. Thereafter, the established relation must also be retained for future implementation. These components are necessary to follow instructions in a manner that enables successful behaviour at the appropriate time. These will be highlighted on below, albeit for the first step of language comprehension.

Memory Representation

To successfully implement behaviour as described by instructions, they must be retained for future implementation. That is, the specified elements must be remembered, such as, remembering which ingredients to retrieve from the refrigerator and which tools to retrieve in the kitchen. Additionally, the instructions may describe the particular circumstances to do some behaviour, like adding ingredients when the mixture turns a certain colour. Failing to remember the elements of the instruction will lead to a partial or full failure to implement the correct behaviour in the situation (e.g., Marcovitch et al., 2010; Roberts & Anderson, 2014). It is generally believed that the working memory is responsible for holding a limited amount of information in a highly accessible state to be used in the short-term (e.g., Baddeley, 1983; Cowan, 1988; Oberauer, 2009). These representations enable the successful retention of the instructed elements such that the correct behaviour can be implemented in correct situations. For instance, in the cooking example above, the person had to add some ingredients when the mixture turned yellow. The person would have to remember that when the mixture turned yellow, the other ingredients had to be added. If the person forgot the critical situation (i.e., yellow), the appropriate response would not occur. Likewise, had the person forgot the appropriate response (i.e., adding the other ingredients), then the appropriate response would not be implemented, even though the person may recall that “something had to be done”. Thus, remembering the elements of the instructions is a critical component of following instructions.

Configuration

In addition to remembering the elements, if the relation between the behaviour and situation has not been specifically implemented, they must be established. For instance, in the cooking example, the person was familiar with both the ingredients and the stirring, but the person had never stirred those specific ingredients before. Because of a lacking relation between those ingredients and the stirring, the person had to configure or connect the ingredients with the stirring behaviour. Without such a relation, the behaviour might not be successfully implemented. Indeed, research suggests that people can understand instructions, in the sense that they can reiterate the instructions, but nevertheless fail to implement them (Drewe, 1974; Luria, 1973; Stuss et al., 2000). This is known as goal-neglect (Duncan et al., 1996) and suggests that merely understanding instructions may not, necessarily, be enough to successfully implement them (e.g., incomplete task model; Duncan et al., 2008). The established relation must also be retained for future implementation in addition to the individual components for a successful implementation of the behaviour in the correct situation.

Preparation

The established relation and the individual elements described by the instructions can vary in their degree of influence. Some representations may entertain a high degree of preparation (or accessibility) that can initiate the behaviour faster than if the representations entertained a lower degree of preparation (or accessibility). The ability to initiate more preparation for an upcoming event is typically referred to as proactive control (Braver, 2012; Braver et al., 2009). Initiating a high degree of preparation creates a state of optimized processing for the relevant perceptual features and prepares the appropriate motor response (Braver, 2012). For example, in the cooking example, the person may focus intently (i.e., high preparation) on the colour of the mixture and be ready to put the ingredients in quickly. If the person had been highly prepared for the behaviour, the person might initiate the behaviour earlier – put the ingredients in too early. On the other hand, if the person was less prepared (or focused), the person might not initiate the behaviour in time – put in the ingredients too late. Thus, the degree of preparation for the instructed relation can influence the behaviour, but not necessarily result in the same failure as the former components do.

Instruction-Based Research

Most research provide participants with instructions about the experiment and the task. Despite such instructions, the interest of researchers has rarely been to investigate the effects of the instructions themselves. Rather, researchers typically provide participants with practice on the task before the main part of the experiment starts. Alternatively, the research may consider the first couple of trials in the (main) experiment as practice and ignore them in the analysis. These steps are often meant to reduce the noise in the data stemming from the learning processes. However, these steps reduce or ignore the important mechanisms of rapidly adapting to the instructions. Additionally, it has been unclear how to investigate the effect directly following the instructions, since only a couple of trials may be sufficient for learning to take place. Only recently developed experimental designs have allowed for a more systematic way to investigate the effects directly following the instructions. These experimental designs have been pivotal in establishing the instruction-based research field that is interested in the adaptive mechanisms that take place before any direct experience of novel instructed relations.

Instruction-based research designs typically rely on the congruency effect (Kiesel et al., 2010; Kornblum et al., 1990), which provides a method to investigate cognitive mechanisms. The congruency effect is typically based on the overlap between two tasks that create conditions that are either in agreement or in conflict. The overlap between the tasks is typically created in relation to some stimulus features, which are specifically related to a response. Depending on the presentation of the stimulus features, they can either be associated with the same response or the opposite response. If the stimulus features are associated with the same response, then responses are typically faster and more accurate than if the stimulus features are associated with the opposite response. The former condition is called the congruent or compatible condition, while the latter is typically called the incongruent or incompatible condition. Since these conditions induce costs and benefits for the behaviour depending on the conditions, experiments can rely on them to measure the degree of influence one task induces on the other.

One important feature of these designs is that one of the tasks presents simple stimulus-response mappings that are not executed until later in the experiment. Consequently, the instructed mappings must be remembered for a later implementation and are thereafter changed. Since the mappings are continually switched and only executed once, practice with the specific mappings is kept low or absent. The overlap between the two tasks creates a congruency effect that enables faster or better performance on congruent trials compared to incongruent trails (Liefooghe et al., 2012; Meiran et al., 2015). Because the mappings are supposed to be novel with, at best, a low degree of exposure, the effect between the tasks must then be related to mechanisms that rapidly adapt behaviour to the instructed relation.

One of the classical designs in the instruction-based research field is the diagnostic-inducer task (Liefooghe et al., 2012). In this design, participants must respond to a single letter with one of the two tasks, depending on the colour of the letter. One of the tasks relates to categorizing the letter as either appearing in italic or upright font (e.g., if italic press left, if upright press right). This task is called the diagnostic task as it diagnoses the effect that the retained instruction has on it. The retained instruction is called the inducer task, as it induces an effect on the diagnostic task. The inducer task changes throughout the experiment and presents participants with two new letters and relates them to a left and right response (e.g., if A press left, if B press right). Thus, participants are first presented with the inducer task, and must then remember the inducer instruction for a later implementation while engaging in the diagnostic task. The overlap between the tasks creates congruent (e.g., italic A) and incongruent (e.g., upright A) trials. Results indicated that a congruency effect is present on the diagnostic task immediately following the inducer instructions, and the authors called this the instruction-based congruency effect. This design has opened questions (Corneille & Béna, 2023) and opportunities for future research to investigate the ability to quickly adapt behaviour to novel relations.

Task Preparation

Investigating how preparation for the inducer task influences the diagnostic task offers a concrete way to investigate the adaptive mechanisms of novel instructed relations. As previously noted, the instructions must be interpretable because no behaviour can successfully be adapted (e.g., interpreting a Chinese recipe when you only know English is unlikely to result in success). In addition, it is crucial that the instructed elements and relations are remembered for future implementation. The relative failure that these two components (memory and relation establishment) create on performances makes them less desirable to use for investigating the adaptive mechanism of novel instructions. Meanwhile, preparation for the inducer task does not necessarily lead to the same failure in behaviour and offers a concrete method to investigate performance. Manipulating the degree of preparation can be done in various ways, as will be highlighted below, without resulting in the failure to successfully implement the instructed relations. Research investigates preparation for the inducer by assuming that a higher degree of preparation induces a greater congruency effect on the diagnostic trials, while a lower degree of preparation results in a smaller congruency effect (i.e., smaller difference between the congruent and incongruent conditions).

Liefooghe et al. (2013) investigated how manipulating two aspects related to preparation influenced the congruency effect on the diagnostic task. In their first experiment, they investigated how repeating the inducer instructions right before the inducer trial would influence the congruency effect on the diagnostic task. Participants had to read the inducer instructions, and after they received a cue indicating whether the inducer instructions would be repeated or not. Results indicated that when the inducer instructions were repeated before the inducer task, participants did not indicate a congruency effect on the diagnostic task. However, when the inducer instructions were not repeated, the participants did indicate a congruency effect on the diagnostic trials. The pattern suggests that when the inducer instructions were repeated, participants could attenuate their degree of preparation for the inducer trial. In their second experiment, they investigated the effect of restricting the response deadline on the inducer trial. Participants received either 1 second (it was actually 2 seconds because it was hard for participants to respond within 1 second) or 5 seconds to response to the inducer. Results indicated that the congruency effect was only present for the 1-second condition, but not the 5-second condition. Thus, participants appeared to attenuate their degree of preparation for the inducer trial depending on the expected utility of preparing for the inducer trial.

A correlational study (Braem et al., 2019) corroborated the link relationship between preparing for the inducer task and the congruency effect. The authors investigated the correlation between the inducer response time and the congruency effect on the diagnostic task. The idea was that if participants were more prepared for the inducer task, they would show faster response times to the trial, which would, in turn, induce a greater impact on the diagnostic task. In line with this idea, the results indicated a negative correlation between the inducer response time and the congruency effect. That is, participants who responded quickly to the inducer trial indicated a greater congruency effect, while participants who responded slower indicated a weaker congruency effect. In addition, the authors investigated whether a similar correlation could be found for a practised version of the task. Intriguingly, they did not find a correlation between the response times on the inducer trial and the congruency effect. The primary difference between the instruction-based congruency effect and the practice-based congruency effect is the memory representation they rely on. The former is assumed to rely on a working memory representation, while the latter is assumed to rely on a long-term memory representation. In sum, the results suggest that preparation may be more influential for novel instructed task-rules, relying on a working memory representation as compared to a practised task.

González-García et al. (2020) investigated how selecting either four or two inducer mappings influenced the congruency effect using an altered version of the diagnostic-inducer task. They presented participants with four different stimulus-response mappings, and a short delay after reading the instructions, they received a retro-cue (Souza & Oberauer, 2016) informing them which of the mappings would be relevant for the current run. The retro-cue could select either all four or two of the instructed mappings. Results indicated that only runs in which two of the instructions were selected induced a congruency effect, while all four instructions did not result in a congruency effect. The result suggests that only a limited number of task-rules may receive a high degree of preparation that, in turn, induce an impact on the diagnostic task. The authors argue that when a limited number of instructions are selected, they engage proactive control that improves processing of the relevant features and prepares the relevant motor response(s) (but see Pereg & Meiran, 2019).

One study (Whitehead & Egner, 2018) investigated whether implicit learning could influence the degree of preparation for the inducer trial. In this experiment, participants engaged in an adapted version of the diagnostic-inducer task, where they had to either recall or implement the inducer instructions. During the recall conditions, participants had to evaluate whether the presented instructions were the same as the instructions presented at the start of the run. During the implementation condition, participants had to respond according to the instructions. Unbeknownst to the participants, the researchers manipulated the proportion of recall and implement conditions, such that some participants mostly received the implementation condition, and others mostly received the recall condition. The results indicate that the congruency effect was modulated in relation to the expected prospective use of the task. Thus, participants who received mostly the implementation conditions indicated a stronger congruency effect, while those who received mostly the recall condition indicated a weaker but still significant congruency effect. The outcome suggests that the expected utility of preparing for the inducer trial influences the congruency effect, presumably because participants are more or less prepared to respond according to the inducer instructions.

Cue Preparation

Task-switching experiment has been extensively studied in the last couple of decades (e.g., Kiesel et al., 2010; Koch & Kiesel, 2022). These designs present participants with two tasks that must either be switched (switch-trial) between or not (repeat-trial). The performance is then compared between the switch trials and the repetition trials. Typically, these designs indicate that switching result in increased response times relative to a repetition trials, which has been termed the switch cost (e.g., Biederman, 1972; Rogers & Monsell, 1995). One way to reduce (Koch & Kiesel, 2022) or abolish (Verbruggen et al., 2007) the switch cost is to present a cue before the upcoming task. Cues have often been combined with task switching designs because it allows for the investigation of preparatory processes. A cue can take many shapes but is typically a symbol that is distinct from other elements of the experiment. These cues inform participants about the upcoming demand such that participants can prepare for the upcoming task (e.g., proactive control; Braver, 2012; Braver et al., 2009).

For example, (Meiran, 1996) investigated how cues influenced the switch cost in a dual task setup. Participants were presented with a 2 × 2 grid that presented a stimulus in one of the four squares. Before the stimulus appeared, a cue was presented indicating which of the two tasks would be relevant for the upcoming stimulus. In one of the tasks, participants had to evaluate whether the stimulus appeared in a left or right horizontal location, irrespective of the up/down location. In the other task, participants had to evaluate whether the stimulus appeared in a vertical location (i.e., top vs. bottom), irrespective of the horizontal location. Results indicated that switching trials had slower response times relative to repetition trials. Additionally, trials that had a longer interval between the cue and the stimulus (cue-stimulus interval) resulted in reduced switch cost relative to a short cue-stimulus interval. The fact that longer cue-stimulus intervals resulted in a reduced switch cost suggests that preparatory processes require some time to be (fully) initiated.

To reduce the switch cost, specific knowledge about the upcoming task is important. Research suggests that switching-to cues reduce the switch cost more than general “switch-away-from” cues. Nicholson et al. (2006) investigated how switch information influenced preparation for task switches. Participants were informed about three different categorization tasks: a letter, a digit, and a colour task. The task related each of the three tasks to two different slices of a 6-cut circle, each task was immediately adjacent to itself. Before the presentation of a stimulus, a cue would indicate which segments, and hence which task(s), were relevant for the upcoming trial. This design made it possible to present participants with three different conditions: Repeat (i.e., the same segment as just responded to), switch-to (i.e., a new segment related to only one task), and switch-away (i.e., a segment related to two other tasks than the one just responded to). In the latter, switch-away case, specific information regarding the upcoming task was unclear, whereas the switch-to condition clearly informed participants about the upcoming task. The results indicated that switching to a specific task (switch-to) resulted in a reduced switch cost relative to the switch-away cue. These results suggest that specific knowledge about the upcoming task is important to initiate preparation for the task, whereas simply “dropping” a task cannot initiate such preparation.

Additionally, the predictability of the cue is another critical aspect that can influence performance. One study (Wendt et al., 2012) investigated the effect of cue validity on performance. These authors presented participants with three different tasks and before each task, they were presented with a cue that indicated the upcoming task with 75 % probability. The results indicated that when the cue did not predict the upcoming task (invalid cue), response times were increased compared to when the cue predicted the upcoming task (valid cue). They also found that the relative increase in response time was if an invalid cue preceded an invalid cue. The authors argue that participants appear to dynamically update the value of relying on the cue if they were recently misguided by them. Moreover, performance was particularly impaired for the task that they had erroneously prepared for, suggesting some lingering inhibition of the task or a reluctance to prepare for the previously prepared for task. In sum, the results suggest that cues need to be considered valid for preparation to be (fully) initiated.

The Present Research

The ability of humans to quickly adapt to novel instructed relations is in part explained by the degree of preparation for the upcoming task. In many circumstances, preparation for an upcoming task is not necessary to do immediately. Rather, preparation for an upcoming task may be initiated at a later point, to a more immediate signal about the upcoming demand. In the present study, I investigated whether preparation for the upcoming task can be delayed until a later point using an explicit signal about the upcoming task demand. To achieve this, I utilized the diagnostic-inducer task. I started by replicating the basic instruction-based congruency effect in experiment 1. Then I proceeded to investigate the effect of presenting participants with a cue 0 to 4 trials before the inducer trial was presented. Thus, preparation for the inducer trial may be attenuated before the cue (pre-cue), resulting in a reduced or absent congruency effect, while after the cue (post-cue) preparation for the inducer task can be initiated, likely resulting in a congruency effect on the diagnostic task.

Experiment 1: Replicating the Instruction-Based Congruency Effect

Methods

Participants

I recruited 33 participants aged 18 to 50 with normal or corrected-to-normal vision from the United Kingdom via Prolific. For the recruitment, I used Prolific's default sampling option (standard sampling), which allows researchers to access a broad and diverse participant pool without applying specific demographic filters beyond those already mentioned. Due to a coding mistake, data from four individuals were not recorded, and the remaining 29 participants had an average age of 33.79 (SD = 8.16; 15 females). Furthermore, 2 participants were excluded due to an overall error rate of more than 30 %. However, these individuals could not be excluded due to lacking connections to the data entries. Participants were paid £3, according to the standard rate (£9 per hour), for an expected time of 20 minutes. All participants were informed that all responses to the task would be recorded and stored for scientific purposes. Furthermore, they were informed that the data would be made publicly available, but would be stored anonymously, such that the data could not be traced back to them. Lastly, participants were informed that they could withdraw their consent at any time by closing the browser tab, and that no data would be stored in that case.

Design

The experiment 1 was a replication of the basic diagnostic-inducer task (Liefooghe et al., 2012, 2013). The experiment consisted of a within-subject design with congruency as the factor. The congruency related to two conditions which were defined based on the overlap between the instructed stimulus-response mappings of the diagnostic task (e.g., if italic press left; if upright press right) and the inducer task (e.g., if abc press left; if def press right). I recorded response time for all trials and errors for the diagnostic and inducer trials. The experiment was preregistered with task material, hypothesis, and the analysis script at the Open Science Foundation (OSF; https://osf.io/fu2w4). Lastly, all material and data are publicly available in the main repository at OSF: https://osf.io/tvkgp/.

Material

A list of 128 three-letter words were randomly sampled from the Subtlex-UK (all) database (van Heuven et al., 2014). Three-letter words were chosen to reduce the chance of key-side association (e.g., A is located on the left side of the keyboard and might be associated with the left hand and thus a left response). Only three-letter words that had a low occurrence frequency (<10) were selected. Each word was tested using Levenshtein distance to ensure that each word was dissimilar by at least 2 characters (e.g., if eug existed in the list, and eua was randomly sampled, it would not be selected due to an overlap in “e” and “u”). This was done to ensure that each combination of non-words was mostly dissimilar.

Each participant received 31 pairs of 3-letter non-words randomly sampled from the generated 128 non-word list. A single 3-letter non-words were presented in the general format “if *3-letter non-word* press *response-side*”. The diagnostic and inducer instructions always appeared in pairs, describing the upright and italic (e.g., if upright press left, if italic press right), or the first and second 3-letter non-word (e.g., if eug press left, if pxv press right – where the stimulus appeared coloured) one above the other. The appearance of the left and right location was randomized between participants, such that some participants always received the “right” response at the topmost row, while others always received a “left” response at the topmost row (see Appendix A inducer instructions for an example). One of the sampled pairs was used for the diagnostic practice, where 16 trials were presented with an equal amount of upright and italic stimuli. Six of the pairs were used for the inducer practice, where each pair were presented in one inducer instructions followed by a single inducer trial. Across the 6 inducer practice trials, an equal distribution of right and left correct responses was presented. A total of 240 diagnostic trials were spread across the remaining 24 pairs of non-words. An experimental block presented one pair of non-words starting with the inducer instructions, relating the two 3-letter non-words to a left and right response. Subsequently, a diagnostic run with a length between 4 and 16 trials were presented, ending with a single inducer trial. Participants responded to the trials using their keyboard, where a left response corresponded to the F-key, and a right response corresponded to the J-key.

The task was coded in JavaScript using the jsPsych library (version 7.3; Leeuw et al., 2023) to be able to run in participant's web browser. The experiment used a light-grey background, but for the feedback questionnaire at the end of the experiment, where the background was set to white. All text appeared in black colour but for the inducer stimuli which appeared coloured (yellow, blue, or red) within the instructions and on the inducer trials. Furthermore, all stimuli and instructions appeared centred on the screen using Open

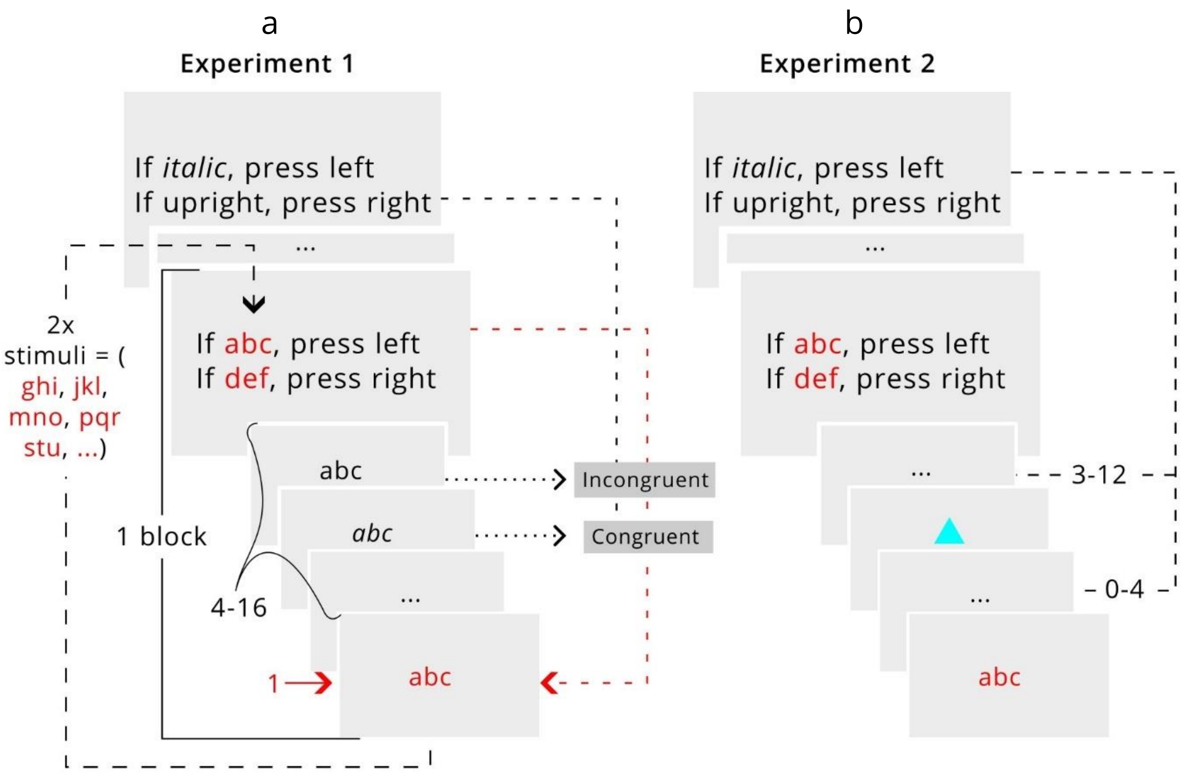
Sans font. Instructions were presented in 24 px, each stimulus was presented in 42 px, and fixation crosses were presented in 48 px.

Procedure

Participants were tested on their own personal computer in a web browser of their choosing, except for Safari (due to incompatibility with full screen). They were informed that the experiment would proceed rapidly without breaks and asked to be in a quiet place where disruptions were unlikely. Participants were informed that they would receive two instructions: one instruction would remain the same throughout the task and one changed through the task. The colour of the appearing 3-letter non-word indicated which of the task participants had to follow. After the task explanation, the screen remained blank for 1500 ms before the diagnostic instructions were presented. Thereafter, participants engaged in the diagnostic practice, followed by the inducer practice instructions and the inducer practice round. Both the diagnostic instructions and the inducer instructions were presented for a maximum of 20 seconds, or until participants pressed the space bar. The diagnostic and inducer trials remained on screen until participants responded with a left or right response. All wrong responses were immediately followed by the screen turning red and displaying "Wrong!" for 300ms. After both the diagnostic and inducer instruction screen, and all trials (diagnostic and inducer) – or after the feedback – were followed by a fixation cross of 750 ms. Before starting the main task, participants were informed that the task would impose a deadline of 2 seconds. Thereafter, 24 blocks of the full procedure were presented: each block started with the inducer instructions, followed by a diagnostic run between 4 and 16 trials, ending with a single inducer trial. All diagnostic and inducer trials were presented for a maximum of 2000 ms or until participants responded. The inducer instructions remained on screen for a maximum of 20 seconds, or until participants press the space bar. All wrong and slow (i.e., more than 2000 ms) trials were immediately followed by the screen turning red and displaying “Wrong!” or “Slow” (respectively) for 300 ms. After every instruction screen and all every diagnostic trial – after an eventual feedback screen – a fixation cross of 750 ms was presented. The inducer trial, however, followed by a fixation cross of 1500 ms, marking the end of the block (Figure 1a). Finishing the 24 blocks presented participants with two open feedback questions: (1) what strategy they used to solve the tasks and (2) optional feedback regarding the experiment in general. Submitting the last feedback redirected participants to Prolific with the completion code.

Figure 1

Overview of the Diagnostic-Inducer Task



Note. The diagnostic instruction (uppermost block) is only shown once. Each block starts with the presentation of a new inducer instruction (instructions with the red three-letter word).

Statistical Analysis

Preregistered. I did two paired-sample t-test for the aggregated dependent variable's response time (RT) and proportion of error (PE) during the diagnostic trials, with congruency as the experimental factor. Moreover, I included two Bayesian paired sample *t*-test with the same predictor and dependent variables. I followed the exclusion criteria used by Liefooghe et al. (2013): Participants with an overall error rate of more than 30 % were excluded from the analysis and the first block was considered a practice round and excluded from the analysis. Furthermore, before aggregating, response times deviating more than 2.5 standard deviations from each individual's mean were excluded from the analysis. Non-responses on the diagnostic trials were excluded and blocks with a wrong response on the inducer trial were excluded from the analysis.

Exploratory. Speed and accuracy might be considered as part of the same behaviour but relating to different outcomes. Because of the close relationship between speed and accuracy, integrated measures of speed and accuracy have been proposed. The application of such integrated measures may be particularly important for tasks relying on both speed and accuracy to measure performance. The present experiment may benefit greatly from using such transformations due to the relevance of both speed and accuracy on measuring performance. In particular, different individuals may adopt different strategies relating to speed and accuracy, possibly leading to null findings or contradictory findings. There is already some research that has used such integrated measures (e.g., Abrahamse et al., 2022; Liefooghe & De Houwer, 2018; Liefooghe & Verbruggen, 2019), but the use is not widespread. I report both the linear integrated speed-accuracy score (LISAS; Vandierendonck, 2017) and the balanced integrated score (BIS; Liesefeld & Janczyk, 2019; see also Liesefeld & Janczyk, 2023). Previous research suggests that the LISAS is slightly biased towards response time, while the BIS returns a balanced integrated score, not biased towards either response time or error (Liesefeld & Janczyk, 2023). For this reason, the BIS may be preferred in the situations where an unbiased integration is required. For both measures, I did a paired sample *t*-test including a Bayesian paired sample *t*-test with the integrated scores as the dependent variable and congruency as the predictor.

To do my analyses, I used R (version 4.3.3; R Core Team, 2024) with the tidyverse (Wickham et al., 2019) for general data tidying and visualization, patchwork (Pedersen, 2024) to combine plots, BayesFactor (Morey et al., 2024) and bayestestR (Makowski et al., 2019) to do the Bayesian analysis. Furthermore, I used lsr (Navarro, 2021) to calculate Cohen's *d*, gt (Iannone et al., 2023) to create tables, jsonlite (Ooms et al., 2023) to extract JSON formatted strings, and boot (Canty et al., 2024) to bootstrap the data.

Results

Preregistered

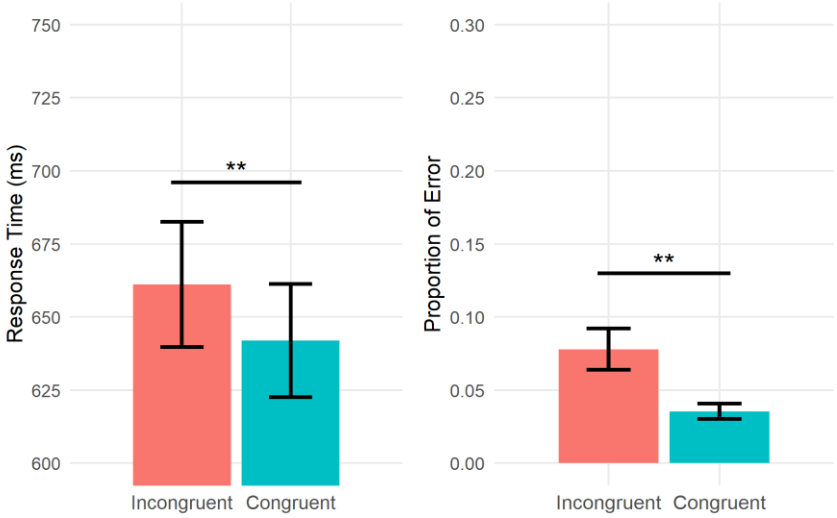
The exclusion criteria resulted in the removal of 30 % (1943 trials) of the data. The first inducer round was considered practice and resulted in the loss of 4.63 % (300 trials). Response times deviating more than 2.5 *SD* and non-responses for 3.19 % (207 trials), and a wrong response on the inducer trial resulted in a loss of 22.2 % (1436 trials) of the diagnostic data (Figure 2). The paired-sample *t*-test for response time, revealed a significant difference (M = 19.2, $t(26)$ = 2.41, p = 012, BF_{10} = 2.31, Cohen's d = 0.46) between the incongruent (M = 661, SD = 111) and the congruent (M = 642, SD = 101) condition. The respective test for the proportion of error revealed a significant difference (M = 0.04, $t(26)$ = 3.17, p = 002, BF_{10} = 10.4, Cohen's d = 0.61) between the incongruent (M = 0.08, SD = 0.07) and the congruent (M = 0.04, SD = 0.03) condition (despite some outliers, removing them did not result in a significant change to the pattern reported, see Appendix B).

Exploratory

The paired-sample *t*-test for the LISAS, revealed a significant difference (M = 50.7, $t(26)$ = 3.43, p = 001, BF_{10} = 18.2, Cohen's d = 0.66) between the incongruent (M = 718, SD = 149) and the congruent (M = 668, SD = 112) condition. The respective test for the BIS revealed a significant difference (*M* = -0.29, *t*(26) = -3.5, $p < .001$, $BF_{10} = 21.5$, Cohen's *d* = 0.67) between the incongruent (*M* = -0.23, *SD* = 0.77) and the congruent (*M* = 0.07, *SD* = 0.57) condition (see Appendix C for a comparison).

Figure 2

Congruency Effect for Response Time and Error Rate



** $p < .01$

Discussion

As hypothesized, the results indicated an instruction-based congruency effect on the aggregated response times and proportion of error, replicating previous findings. Similarly, both the LISAS and the BIS indicated a significant instruction-based congruency effect. The BIS indicated a slightly higher effect size compared to the LISAS, which is related to the way the LISAS integrates response times and errors. As discussed elsewhere (Liesefeld & Janczyk, 2023), the LISAS weighs response time relatively more than errors compared to BIS. The BIS, therefore, may be preferred in situations where effects on errors are typical.

Experiment 2: Preparation and Cues

In this experiment, I aimed to investigate whether preparation for the inducer task could be adjusted to an explicit cue about the upcoming demand. Research has indicated that a high degree of preparation (e.g., Liefooghe et al. 2013, experiment 2) is a component for the congruency effect to appear. Preparation, however, can be delayed to a later point using explicit cues. For example, while waiting for a red light to turn green, one is unlikely to prepare to accelerate until the traffic light turns yellow. Similarly, preparing for the inducer task might be delayed until the explicit cue indicates that the task will soon appear. I hypothesized that participants would indicate a smaller or absent congruency effect before the cue but show a congruency effect after the cue. That is, I expected participants to only indicate fewer errors (smaller LISAS, higher BIS) during the congruent and more errors (larger LISAS, smaller BIS) during the incongruent condition after the cue, but not before the cue. I decided not to include response time in the hypothesis because of lacking power to detect an effect with the reduced number of trials expected after the cue (see Appendix D).

Methods

Participants

I recruited 38 participants aged 18 to 50 with normal or corrected-to-normal vision from the United Kingdom via Prolific. For the recruitment, I used the same recruitment parameters as in experiment 1 in addition to excluding participants who partook in the first experiment. Participants had an average age of 33.79 (SD = 8.16; 15 females). Three participants were excluded because of an overall accuracy of more than 30 %, leaving me with 35 participants. Note that I preregistered 34 participants, however, due to a time-out issue in the recruitment process, one additional participant was recruited. Moreover, due to lacking data-participant link, I cannot be certain which participant to remove. Therefore, I report the analyses of the full sample (35) and include robustness checks of the results by analysing all combinations of the 34 participants. For the analyses of the 34 participants, I report the minimum and maximum values of the tests, indicating the range of possible values the sample could take, had one participant been excluded. Participants were paid £3, according to the standard rate (£9 per hour), for an expected time of 20 minutes. All participants were informed that all responses to the task would be recorded and stored for scientific purposes. Furthermore, they were informed that the data would be made publicly available, but would be stored anonymously, such that the data could not be traced back to them. Lastly, participants were informed that they could withdraw their consent at any time by closing the browser tab, and that no data would be stored in that case.

Design

The experiment consisted of a within-subject design with Congruency (congruent vs. incongruent) and Cue (pre- vs. post-cue) as the experimental factors. The congruency related to two conditions which were defined based on the overlap between the instructed stimulus-response mappings of the diagnostic task (e.g., if italic press left; if upright press right) and the inducer task (e.g., if abc press left; if def press right). The cue split the diagnostic run in two parts: One before the cue (pre-cue) and one after the cue (post-cue). I recorded response time for all trials and errors for the diagnostic and inducer trials. The experiment was preregistered with task material, hypothesis, and (a partial) analyses script at the OSF (https://osf.io/9ejyx).

Material

I used the same task created in experiment 1 with a couple of changes. A cue was introduced to the experiment, as either a square, circle, or triangle presented within a 70 px square (diameter). The cue appeared coloured as one of the two remaining colours after the inducer colour had been selected. An additional round of practice was added, increasing the total amount of 3-letter non-words for each participant to 32. The additional practice round consisted of 13 pre-cue and 3 post-cue diagnostic trials were presented during the added practice round. For the experiment itself, 240 diagnostic trials were presented, split between the pre-cue and post-cue diagnostic run. The pre-cue diagnostic run related to diagnostic trials presented before the cue, while the post-cue related to diagnostic trials after the cue. There were presented in total 168 diagnostic (70 %) trials during the pre-cue run and 72 diagnostic (30 %) trials in the post-cue run. Moreover, I forced an equal number of congruent and incongruent trials across the pre- and post-cue run. This was done to ensure sufficient power for the congruency effect for the short post-cue diagnostic run (as estimated in Appendix D). An additional instruction screen was added to explain the integrated practice round and the other instructions were slightly changed to improve clarity (Appendix E).

Procedure

An additional practice round was added, that presented the structure of all subsequent rounds: inducer instructions, diagnostic trials, cue, and inducer trials. First, the inducer instructions were presented, then the pre-cue diagnostic run presented 13 trials with an equal number of left and right responses before the cue appeared. The cue remained on screen for 1250 ms before the post-cue diagnostic run of 3 trials were presented, after which a single inducer trial appeared. As with the previous practice rounds, none of the trials had a response deadline. The subsequent blocks had the same format but had a random pre-cue length between 3 and 12 diagnostic trials, and a post-cue run between 0 and 4 diagnostic trials (Figure 1b).

Statistical analysis

Preregistered. I used the same exclusion criteria as in experiment 1. I did three repeated measures analysis of variance (rmANOVA) on the aggregated dependent variable proportion of error (PE), and the calculated variables LISAS and the BIS. As for the predictors, I used Congruency (incongruent vs. congruent), Cue (pre-cue vs post-cue) and their interaction Congruency × Cue.

For the Bayesian model, I used a Bayesian linear mixed model implemented via the brms-package (Bürkner, 2017). The predictors were Congruency, Cue, and their interaction for the dependent variables error rate, LISAS and BIS. The brms package relies on the Stan software (Stan Development Team, 2023; via package cmdstanr; Gabry et al., 2023) that implements a Hamiltonian Monte-Carlo algorithm. I used 6 parallel chains with 6000 samples each and used the standard warm up of half the sample (i.e., 3000), leaving us with 3000 samples. Convergence of all models was confirmed visually and all $\hat{R}$-values were confirmed to be lower than 1.05. As for the priors, I used the default priors implemented in the brms that are non-informative for coefficients corresponding to fixed effects and weakly informative for the intercept and standard-deviation parameters (Student-t prior with 3 df, mean = 0 and SD = 2.5). I report the posterior mean b, 95% highest-density interval (HDI), and the probability that the effect is in the specified (i.e., b) direction (p_b). Lastly, I report the evidence ratio (ER_b) that quantifies how much more likely the effect is to be in the specified (b) direction versus the other (e.g., b = -1, p_b = .95, ER_b = 12.1, indicates strong evidence for a negative b).

Exploratory. I included a rmANOVA and a Bayesian linear mixed model for the aggregated variable response time with Congruency, Cue, and their interaction as the predictors. This analysis was included for exploratory purposes and was not used to inform the hypotheses. Moreover, I included an independent sample *t*-test of the aggregated response time on the inducer trial between the first and second experiment. This was included to investigate whether participants were more prepared for the inducer trial in the second experiment compared to the first experiment. Lastly, I included an exploratory analysis investigating the effect of Time. Time was defined as the block count (one block from the inducer instructions to the inducer trial, see Figure 1a), such that the Time variable related to 23 blocks. I scaled the Time variable to range from 0 to 1 such that the coefficient related to Time indicate the change in the coefficients from the start to the end of the experiment. One linear mixed model and one Bayesian linear mixed model were done for each of the four dependent variables (response time, error rate, LISAS and BIS) with all individual predictors (i.e., Time, Congruency, Cue), all two-way interactions (e.g., Congruency × Time), and the three-way interaction (i.e., Congruency × Cue × Time).

To do my analyses, I used R (version 4.3.3; R Core Team, 2024) with the tidyverse (Wickham et al., 2019) for general data tidying and visualization, broom (Robinson et al., 2023) for analyses tidying, gt (Iannone et al., 2023) to create tables, rstatix (Kassambara, 2023) to get outliers, emmeans (Lenth et al., 2024) to estimate the rmANOVA coefficients and pbapply (Solymos et al., 2023) for a progression bar for the apply functions. To do the statistical tests, I used afex (Singmann et al., 2024) for the rmANOVA, and lme4 (Bates et al., 2023) to get the estimated coefficients, brms (Bürkner, 2017) for the Bayesian linear regression models, cmdstanr (Gabry et al., 2023) to run the Bayesian models (connected to Stan; Stan Development Team, 2023), bayestestR (Makowski et al., 2019) to get the highest density interval (HDI) and bayesplot (Gabry et al., 2024) for Bayesian diagnostics.

Results

Preregistered

The first inducer round was considered practice and resulted in the loss of 3.7 % (311 trials). Response times deviating more than 2.5 *SD* and non-responses for 3.63 % (305 trials), and a wrong response on the inducer resulted in a loss of 15.6 % of the data (1307 trials). In total, the exclusion criteria resulted in the removal of 22.9 % (1923 trials) of the data.

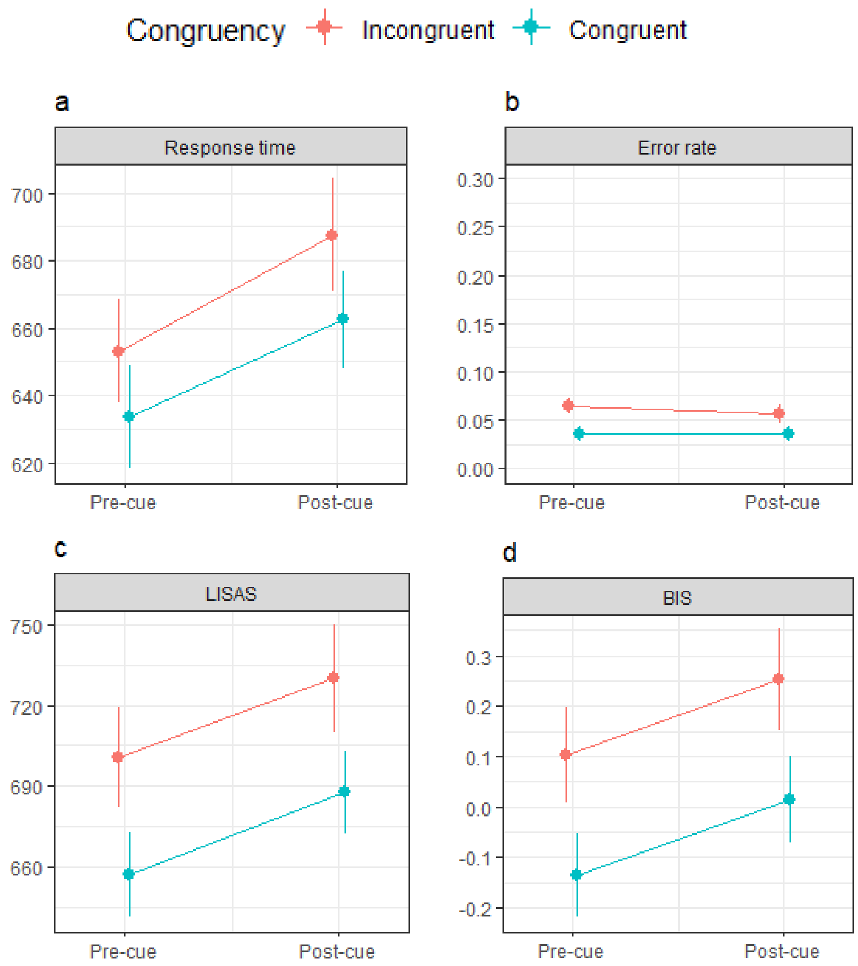
The rmANOVA for the error rate (Figure 3b) revealed a significant effect of Congruency (F(1, 34) = 8.71, p = .006, η_p^2 = 0.204), suggesting lower errors for congruent trials. The results did not reveal a significant main effect or Cue (F(1, 34) = 0.36, p = .552, η_p^2 = 0.01), nor an interaction effect (F(1, 34) = 0.19, p = .669, η_p^2 = 0.005). As for the LISAS (Figure 3c), the results revealed a main effect of Congruency (F(1, 34) = 30.8, p < .001, η_p^2 = 0.476) suggesting lower LISAS on congruent trials, and a significant main effect of Cue (F(1, 34) = 15.1, p < .001, η_p^2 = 0.308), suggesting that the LISAS increased after the cue.

However, no interaction was found (F(1, 34) = 0.01, p = .942, η_p^2 = 0). The rmANOVA for the BIS (Figure 3d, note that the BIS is inversed in the figure to correspond to the other plots) revealed a main effect of Congruency (F(1, 34) = 25.2, p < .001, η_p^2 = 0.426), suggesting higher BIS for congruent trials. However, no main effect of Cue (F(1, 34) = 9.67, p = .004, η_p^2 = 0.221), or an interaction (F(1, 34) < 0.01, p = .995, η_p^2 = 0) was found (see Table F1, F2 and F3 for a comparison). Testing all combinations of a sample size of n = 34 did not significantly change the pattern of results (see Table F4) and neither did removing the outliers (Figure F1 and Table F7).

For the Bayesian models (Table F5) of the error rate, results indicate extreme evidence that a congruent trial reduced the number of errors ($b = -0.03, [-0.05,\ -0.01], p_b = 0.99, \mathrm{ER}_b = 163$). However, the analysis only provided anecdotal evidence for a decrease in errors after the Cue ($b = -0.01, [-0.03, 0.01], p_b = 0.75, \mathrm{ER}_b = 3.00$), and no evidence for an interaction ($b = 0.01, [-0.02, 0.04], p_b = 0.67, \mathrm{ER}_b = 2.00$). In relation to the LISAS, the analysis revealed extreme evidence for a congruent reduction in LISAS ($b = -43.4, [-65.5,\ -21.7], p_b = 1.00, \mathrm{ER}_b = \infty$) and an increase in the LISAS after the Cue ($b = 29.4, [8.41, 51.9], p_b > 0.99,\ \mathrm{ER}_b = 230$). However, no interaction between Congruency and the Cue was found ($b = 0.91, [-29.9, 32.4], p_b = 0.52, \mathrm{ER}_b = 1.09$). Lastly, for the BIS, the analysis revealed extreme evidence for a congruent increase in the BIS ($b = 0.24, [0.11, 0.37], p_b > 0.99, \mathrm{ER}_b = 2570$) and very strong evidence a decreased BIS after the Cue ($b = -0.15, [-0.28,\ -0.02], p_b = 0.99, \mathrm{ER}_b = 80.45$). However, no interaction was observed ($b = 0.00, [-0.18, 0.19], p_b = 0.50, \mathrm{ER}_b = 0.99$). Furthermore, the Bayesian model with n = 34 did not significantly change the pattern of results (Table F6), neither did removing the outliers (Table F8).

Figure 3

The Effect on the Dependent Variable Under the cue and Congruency Conditions



Note. The balanced integration score (BIS; Liesefeld & Janczyk, 2019) is inverted to correspond to the other plots. LISAS = Linear integrated speed-accuracy score (Vandierendonck, 2017).

Exploratory

The rmANOVA for response time (Figure 3a) revealed a main effect of Congruency (F(1, 34) = 30.3, p < .001, η_p^2 = 30.3), suggesting faster response times for congruent trials. As for the cue, the analysis revealed a significant main effect of the Cue (F(1, 34) = 26.3, p < .001, η_p^2 = 26.3), suggesting that response time was slower after the Cue. The analysis did not reveal an interaction between the Congruency and the Cue (F(1, 34) = 0.3, p = .588, η_p^2 =0.3). In relation to the Bayesian mixed model of response time, it revealed extreme evidence for a congruent decrease in response times ($b = -19.5, [-34.3, -4.92], p_b > 0.99, \mathrm{ER}_b = 213$). Additionally, response times were slower after the Cue ($b = 34.5, [20.2, 49.4], p_b = 1.00, \mathrm{ER}_b = \infty$), and no interaction between was observed ($b = -5.73, [-26.4, 15.0], p_b = 0.70, \mathrm{ER}_b = 2.37$).

Comparing response times to the inducer trial between the first (*M* = 923, *SD* = 209) and the second (*M* = 893, *SD* = 170) experiment revealed anecdotal evidence for no difference (M_{diff} = 30.1, $t(49.4) = 0.61$, $p = .273$, $BF_{01} = 3.22$). Lastly, the exploratory analyses of the three-way interaction with time (i.e., Congruency × Cue × Time) only revealed a significant effect for the dependent variable LISAS (Table 1; $t(2293) = -2.45$, p =

Table 1

Coefficients for the Three-way Interaction Model for the LISAS

	Frequentist				Bayesian				
								HDI	
Predictors	β	*df*	*t*	*p*	*b*	p_b	ER_b	Low	High
Congruency	-91.3	2292	-4.76	< .001	-95.1*	1.00	Inf	-135	-55.6
Cue	5.91	2293	0.29	.774	4.06	.570	1.33	-38.7	47.8
Time	-72.2	2294	-3.10	.002	-3.28*	.999	1285	-5.31	-1.23
Congruency × Cue	68.2	2293	2.36	.018	73.8*	.993	133	14.1	135
Congruency × Time	81.8	2292	2.47	.014	3.72*	.994	154	0.79	6.63
Cue × Time	45.6	2293	1.29	.197	2.05	.899	8.88	-1.11	5.14
Congruency × Cue × Time	-122	2293	-2.45	.014	-5.55*	.993	152	-9.94	-1.10
Model fit									
Sigma (subject)					177*	1.00	Inf	172	182
R^2					.248			.221	.274
LOOIC					30831			SE = 193	

Note. The Time (block) variable is scaled to range from 0 to 1. Thus, the three-way interaction coefficient indicates the adjustment of the Congruency × Cue at the end of the experiment in contrast to the start of the experiment. p_b = probability that the effect is in the noted (*b*) direction; ER_b = evidence ratio for the noted (*b*) direction; LISAS = Linear integrated speed-accuracy score (Vandierendonck, 2017).

* p_b > .95

.014, $b = -5.55, [-9.94,\ -1.10], p_b = 0.99, \mathrm{ER}_b = 152$) and BIS (Table 2; $t(2293) = 2.15$, $p = .032$, $b = 0.03, [0, 0.05], p_b = 0.98, \mathrm{ER}_b = 62.6$), but not for response time (Table G1; $t(2293) = -1.35$, $p = .176$, $b = -2.07, [-5.06,\ 0.89], p_b = 0.913, \mathrm{ER}_b = 10.5$) or error rate (Table G2): $t(2298) = -1.7$, $p = .089$, $b = 0, [-0.01,\ 0.00], p_b = 0.95, \mathrm{ER}_b = 19.9$. Note, however, that both response time and error rate receive strong support from the Bayesian model. See Table 1 and 2 (G1 and G2) for an overview of all the coefficients (see Figure G1 and G2 for a visual representation).

Table 2

Coefficients for the Three-way Interaction Model for the BIS

	Frequentist				Bayesian				
								HDI	
Predictors	*β*	*df*	*t*	*p*	*b*	p_b	ER_b	Low	High
Congruency	0.51	2292	4.71	< .001	0.53*	1.00	Inf	0.3	0.76
Cue	-0.04	2293	-0.33	.745	-0.03	.592	1.45	-0.28	0.21
Time	0.39	2294	2.97	.003	0.02*	.999	666	0.01	0.03
Congruency x Cue	-0.35	2293	-2.16	.031	-0.38*	.985	64.7	-0.73	-0.04
Congruency x Time	-0.44	2293	-2.35	.019	-0.02*	.990	102	-0.04	0.00
Cue x Time	-0.19	2293	-0.96	.335	-0.01	.831	4.91	-0.03	0.01
Congruency x Cue x Time	0.61	2293	2.15	.032	0.03*	.984	62.6	0.00	0.05
Model fit									
Sigma (subject)					1.00*	1.00	Inf	0.97	1.03
R^2					.208			.182	.235
LOOIC					6685			SE = 156	

Note. The Time (block) variable is scaled to range from 0 to 1. Thus, the three-way interaction coefficient indicates the adjustment of the Congruency × Cue at the end of the experiment in contrast to the start of the experiment. p_b = probability that the effect is in the noted (*b*) direction; ER_b = evidence ratio for the noted (*b*) direction; BIS = Balanced integration score (Liesefeld & Janczyk, 2019).

* $p_b > .95$

Discussion

Overall, the preregistered analysis did not reveal the expected pattern. Participants indicated a significant congruency effect before and after the cue. However, they did not indicate a significant interaction between the cue and the congruency effect, suggesting that the congruency effect was similar before and after the cue. Although I did not hypothesize any effect on response time, an exploratory analysis revealed that response time was significantly slower after the cue compared to before the cue. This slowdown was not accompanied by any significant changes in error rate, as might be expected if participants changed their speed-accuracy trade-off strategy. The pattern of results suggests that the cue primarily influenced response time, and only by slowing the responses after the cue. One reason for the general slowdown in response time, might have been that participants were more prepared for the upcoming inducer task, and therefore spent longer responding to the diagnostic trials. To investigate this possibility, I tested the response time to the inducer trial between the first and second experiment, but the test did not reveal a difference, albeit numerically faster in the second experiment compared to the former. This result suggests that the slowdown during the diagnostic trials did not significantly influence response time on the inducer trial.

General Discussion

I investigated whether the instruction-based congruency effect could be attenuated by an explicit cue of the upcoming switch. To achieve this, I first replicated the instruction-based congruency effect using the diagnostic-inducer task (Liefooghe et al., 2012). Following a successful replication, I introduced a cue in the diagnostic task between 0 and 4 trials before the inducer trial appeared (i.e., before the inducer trial become relevant). Contrary to my preregistered hypotheses, participants did not indicate a significant two-way interaction between the cue and the congruency, meaning they showed a similar congruency effect before and after the cue. Results from the exploratory analysis suggested that the cue only slowed response after the cue relative to before the cue, but no changes in error rate were found. In light of recent research, proposing that cognitive control might require learning (Abrahamse et al., 2016; Braem et al., 2017, 2024), I explored the influence of time on the relationship between the congruency effect and the cue. Results suggested that time significantly influenced the interaction between the congruency effect and the cue. This interaction suggests that the hypothesized two-way interaction, namely, no congruency effect before the cue and a congruency effect after the cue, was found for the later parts of the experiment, but not the first and middle parts of the experiment (Figure G1 and G2).

For the second experiment, I preregistered the analysis by summarizing the congruency effect and the cue throughout the experiment, as is typically done in the field in the experiment (e.g., Liefooghe et al., 2013; Liefooghe & De Houwer, 2018). I expected the congruency effect to be reduced or absent before the cue compared to after the cue. This attenuation was related to the idea that preparation is a key component of the instruction-based congruency effect, and explicitly signalling the upcoming demand (i.e., when the inducer task starts to become relevant) might attenuate preparation for the task before the signal because the participants can be sure that, before the cue, the instructions are not yet relevant. However, the preregistered analyses did not indicate the expected two-way interaction and suggests that the congruency effect was similar before and after the cue. The non-significant interaction is perhaps not that surprising, given that nothing in the task changed. That is, the inducer task remained the same throughout any block, and given the short length of each block, participants might have retained the same degree of preparation (i.e., a high degree) for the inducer trial, regardless of the cue. In this sense, the cue may not have been utilized by the participants since it did not reveal any new information above and beyond what they already knew – namely, that the inducer trial is “soon-to-appear”. Indeed, the lacking predictability of the cue might be a reason for the failure to adapt preparation to the cue (e.g., Wendt et al., 2012).

Typically, cues are used to directly predict an upcoming demand (i.e., next trial, e.g., Koch, 2001; Rogers & Monsell, 1995), and with sufficient preparation (Meiran, 1996), the switch cost can be reduced (Kiesel et al., 2010) or completely disappear (Verbruggen et al., 2007). In my experiment, however, the cue did not provide precise information about the inducer trial. Participants were explicitly told that the cue indicated that the inducer trial would be presented “within some trials (screens)” (Appendix Eh) after the cue. The uncertainty and vagueness of the cue might have made participants weary of relying on it. This would be in line with research indicating that the predictability of the cue is a key component of initiating preparation for a task switch (Wendt et al., 2012). This predictability, however, was something I wanted to avoid because predictability provided by the cue might delay preparation for the switch to right before the task is to switch (i.e., preparation is initiated after, e.g., 3 trials). That is, if participants knew that the task would always switch after three trials, then preparation could be withheld (i.e., not highly prepared for) until after the last (e.g., third) trial, and therefore no congruency effect would be present after the cue. Thus, I wanted the cue to be pseudo-predictable in the sense that it did predict exactly when the upcoming switch would be present, but that it would appear soon. Moreover, the number of trials after the cue was purposefully limited such that the cue would predict the upcoming task within a reasonable amount of time. Nevertheless, future research may want to investigate the effect of predictability on the congruency effect to elucidate whether such preparation would occur at a later point or would be present immediately after the cue.

Learning

Even though the preregistered analysis did not reveal the expected two-way interaction between the congruency effect and the cue, I explored the influence of time on the two-way interaction (i.e., a three-way interaction). The results indicated a significant three-way interaction, suggesting that time significantly adjusted the interaction between the congruency and the cue. Indeed, on closer inspection of the data, the hypothesized two-way interaction (i.e., no/reduced congruency effect before the cue, and a congruency effect after the cue) was found for the later parts of the experiment, but not the earlier parts of the experiment. The results can be interpreted as a gradual adjustment (i.e., reduction) of preparation for the inducer task, before the cue, over time. Indeed, since participants were not exposed to the inducer instructions again, they would have to retain the inducer instructions to successfully implement the instructions. This contrasts to Liefooghe et al. (2013) experiment 1, where participants could, conceivably, forget the instructions (e.g., Lewis-Peacock et al., 2018), and then read and prepare for the instructions once they appeared again, right before the inducer trial. After the cue, however, the congruency effect was present, suggesting that preparation for the inducer task was initiated or improved (Figure G1 and G2).

This finding is in line with recent theoretical perspectives holding that cognitive control might be based on associative learning mechanisms (Abrahamse et al., 2016; Braem et al., 2024; Braem & Egner, 2018). These theories suggest that events can become associated with each other in relation to their known or assumed relationship. For example, if pressing a button provides a food reward, the button can quickly become associated to the food reward. If other contingent effects occur because of the button pressing (e.g., a light) that event can also become associated with the food reward. In the end, the single event (e.g., light) could elicit a desire to push the button or check the food, due to its association to the two elements. Similarly, the implementation of higher-order cognitive control can be associated with events. For example, one study found implicit learning of the expected implementation of the inducer task adjusted the congruency effect – presumably because of adjustments to preparation for the inducer trial (Whitehead & Egner, 2018). Similarly, Braem et al. (2017) showed that the inducer instructions could become associated to a specific location, such that, over time, only diagnostic trials presented in the location where the inducer instruction were presented, indicated a congruency effect, but not diagnostic trials presented in another location.

Future Research

One principle of associative learning is that multiple features can become bound or associated to the relevant response. Thus, irrelevant features of the task may become bound to the associated adjustment of cognitive control (e.g., reduced preparation). For example, the experiment presented participants with a grey background, which may be an eliciting feature that is necessary for the associated adjustment of cognitive control. Indeed, research suggests that cognitive training programs rarely transfer to other tasks than the one they were trained in (Simons et al., 2016; see also Braem et al., 2024). This is also exemplified in Braem et al.'s (2017) study, where the inducer instructions became associated to a specific location of the task. That is, over time, only diagnostic trials presented in the same location as the inducer instruction indicated a congruency effect, while diagnostic trials presented in a different location did not exhibit the same degree of congruency effect. In my experiment, however, it is unclear whether a similar context dependence is present. It is conceivable that, over time, the inducer instructions became associated with the adjusted degree of preparation or became the cue for the adjusted degree of preparation for the inducer trial. Thus, the inducer instructions could inform the new stimulus-response mapping and indicate a relaxed degree of preparation. On the other hand, the cue itself, could be (or remain) associated with the increased degree of preparation (i.e., cognitive control). Future research may want to investigate contextual factors and whether they become associated with the degree of preparation (e.g., what would happen to the congruency effect if the background colour changed?).

Another avenue for future research could be to investigate the effect of reducing the value of preparing for the inducer task. This could be achieved by introducing uncertainty about the upcoming inducer task. For example, the cue could provide information regarding the upcoming task in such a way that preparing for the instructed mappings may not be the optimal strategy. Under this design, the cue becomes a critical piece of information about the upcoming inducer task, and hence, preparation for the instructed inducer task may be reduced from the onset. As an example, the cue could provide information about whether the inducer task remains the same or reverses as the instructed mapping. Due to the possibility that the instructed mappings may change, participants may not prepare to the same degree, since the upcoming inducer task is uncertain. Thus, only after the appearance of the cue will the task be certain, and (a high degree of) preparation for the inducer trial can be initiated.

The results from my second experiment suggests that cognitive control (i.e., preparation) appears to require learning. Therefore, it is possible that participants may display a similar degree of influence before the cue regardless of the value of preparing for the inducer task. Moreover, given the relatively simple stimulus-response mappings, the cost of initiating a high degree of preparation may be low. The simple nature of the instructions might mean that changes can easily be made to the instructions without compromising on the degree of preparation. Thus, a congruency effect may be present both before and after the cue, even though the mappings themselves may have reversed. One possible solution to this problem could be to increase the complexity of the instructed mappings. The complexity might reduce the degree of preparation for the original instructed mapping from the very onset, both because the mapping is complex and the value of initiating a high degree of preparation is low. With that said, complexity in the instructed mappings might make it difficult to initiate any degree of preparation, which might reduce the congruency effect. Although research suggests that working memory load does not seem to influence the instruction-based congruency effect, at least when the load is multiple stimulus-response mappings rather than complexity (Pereg & Meiran, 2019).

The idea that higher-order cognitive control might require learnings has some important implications for future research. Firstly, summarizing across all or multiple blocks may mask the gradual changes in performance. These changes may be particularly significant for instruction-based research to consider, since learning of the specific instructed stimulus-response mappings does not take place. The lack of practice may provide insightful information on higher order mechanisms, as any behavioural changes may be specific to such higher order mechanisms rather than “lower level” changes between the stimulus and the response. For example, Bugmann and colleagues (2019) proposed that participants may establish a “higher-order function” that takes the instructed stimulus-response mapping as a parameter. Through practice with the task, the implementation of the function is improved rather than the individual stimulus-response mappings. Critically, because instruction-based research presents many stimulus-response mappings that are only executed once, improvements in behaviour can be attributed to such a “higher-order function” rather than the direct improvement in the individual stimulus-response mappings. Furthermore, considering the associative nature of the brain (Abrahamse et al., 2016), the “function” could eventually incorporate surrounding features, such as the situation (e.g., experiment room) and task features (e.g., task background, task presentation format). Thus, only under those peculiar situations (e.g., in the scientific room with the task format and the grey background) will the specific “function” be applied (associated with) – possibly explaining why cognitive training rarely generalize (Simons et al., 2016). Instruction-based research may be particularly well suited to elucidate on such abstract learning mechanisms, as the lower-level associations are kept at a minimum. However, to investigate these mechanisms, it will be important to consider the influence of time, as illustrated here and elsewhere (Braem et al., 2017).

As an example, a recent study investigated whether cancelling the upcoming inducer task would remove the congruency effect (Abrahamse et al., 2022). These authors did not find that a cancellation cue removed the congruency effect unless the inducer instructions were replaced. Since adjustment of cognitive control (i.e., cancelling) might rely on learning, it is possible that the cancellation of the inducer task would only appear later in the experiment. The authors did not test the effect of time, and my investigation of their data did not reveal a significant interaction with time (Appendix H). However, these authors presented participants with different run types, and one of these runs related to immediately responding to the inducer trial, while the other runs presented the cue and then a new diagnostic run. Therefore, the utility of attenuating cognitive control might have been reduced, because of the possibility that the instructions will be implemented. Moreover, the additional run types might have reduced the practice of the task conditions, such that sufficient practice did not occur. Thus, future research may want to investigate whether the cancellation of the inducer task may appear after sufficient practice with a standardized design (i.e., always presenting a cue).

For the second experiment, I decided not to use predictable cues. This was because the predictable nature of the switch would be known, and preparation for the inducer task could be delayed until right before the switch. In contrast, the current implementation meant that the trial right after the cue could be the switch, and so participants should be encouraged to prepare directly following the cue. Nevertheless, it could be intriguing to investigate whether delaying the appearance of the inducer trial to some set number of trials after the cue would lead to a similar delay for the inducer task. Alternatively, explicitly informing participants about the number of trials after the cue, before the inducer trial, could provide further insight into whether the expected reduction in the congruency effect is present, or if it requires practice. Future research implementing these manipulations could provide insight into the learning processes of higher-order cognitive control. Moreover, these manipulations could be informative in relation to whether these learning processes differ between populations (e.g., working memory capacity and intelligence, Hülür et al., 2019; Lin et al., 2022).

The results suggested that the hypothesized two-way interaction was only found in the last third of the experiment. Additionally, participants in my experiments finished the experiments faster (~15 minutes) than I expected (~20 minutes). Future research could increase the length of the experiment and further elucidate on the time-course of the learning processes. One might hypothesize that after a certain point of time, the control parameters are fine-tuned, and no further adjustments will occur. Moreover, if such a learning process takes place, it could be interesting to investigate what happens if the cue is suddenly removed, and the inducer trial is presented. Because of the plausibly reduced preparation before the cue, responses to the inducer trial could be reduced. It might then be that the participants must go through a similar (un-)learning process to initiate a high degree of preparation for the inducer trial immediately following the instructions.

Limitations

Even though the experiments were preregistered, the three-way interaction of time was not. Therefore, the three-way interaction should be considered as tentative evidence until future replications confirm it. Despite being in line with prior research (Braem et al., 2017) and theories (Abrahamse et al., 2016; Braem et al., 2024). Additionally, it may be necessary to replicate the experiment in laboratory settings to control for confounding variables that might arise due to online data collection. Indeed, one participant timed out, suggesting that the participants might have been distracted or busy with other things before starting the experiment. Moreover, the experiments indicated a high loss of data due to the exclusion criteria, even though the diagnostic-inducer task typically results in fairly high loss of data (i.e., >10 %, e.g., Abrahamse et al., 2022; Liefooghe & De Houwer, 2018). The relatively short amount of time participants spent on the task combined with a relatively high loss of data could significantly influence the analysis. Future research may reduce the loss of data by either providing more practice or generally increasing the length of the experiment.

Conclusion

In this study, I aimed investigate whether preparation for an upcoming novel instructed task relation could be delayed until an explicit cue about the upcoming task demand. To achieve this, I investigated how a new task instruction influenced the performance on a secondary task before its implementation (Liefooghe et al., 2012). A cue was presented cue a couple of trials before the task switched to indicate the upcoming demand. I hypothesized that the congruency effect would be reduced before the cue, but not after. The preregistered analysis did not reveal the hypothesized pattern. However, an exploratory analysis suggested that the hypothesized pattern was present for later parts of the experiment (i.e., last third) compared to the earlier parts. I argue that the reduced congruency effect before the cue stems from an attenuation of preparatory processes for the retained instructions. This is in line with previous research (Braem et al., 2017) and theories (Abrahamse et al., 2016; Braem et al., 2024; Braem & Egner, 2018), proposing that cognitive control might rely on associative learning mechanisms requiring practice to be adjusted. Future replications are necessary to firmly establish the pattern of results observed in this exploration, as it was not preregistered.

References

Abrahamse, E., Braem, S., De Houwer, J., & Liefooghe, B. (2022). Tenacious instructions: How to dismantle newly instructed task rules? *Journal of Experimental Psychology: General*, *151*(11), 2812–2832. https://doi.org/10.1037/xge0001233

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

Baddeley, A. D. (1983). Working Memory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *302*(1110), 311–324. http://www.jstor.org/stable/2395996

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., Krivitsky, P. N., & Tanaka, E. (2023). *lme4: Linear Mixed-Effects Models using 'Eigen' and S4* (1.1-35.1) [Computer software]. https://cran.r-project.org/web/packages/lme4/index.html

Braem, S., Chai, M., Held, L. K., & Xu, S. (2024). One cannot simply 'be flexible': Regulating control parameters requires learning. *Current Opinion in Behavioral Sciences*, *55*, 101347. https://doi.org/10.1016/j.cobeha.2023.101347

Braem, S., Deltomme, B., & Liefooghe, B. (2019). The instruction-based congruency effect predicts task execution efficiency: Evidence from inter- and intra-individual differences. *Memory & Cognition*, *47*(8), 1582–1591. https://doi.org/10.3758/s13421-019-00951-3

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

Braem, S., Liefooghe, B., De Houwer, J., Brass, M., & Abrahamse, E. L. (2017). There are limits to the effects of task instructions: Making the automatic effects of task instructions context-specific takes practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(3), 394–403. https://doi.org/10.1037/xlm0000310

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

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*(18), 7351–7356. https://doi.org/10.1073/pnas.0808187106

Bugmann, G., Goslin, J., & Thill, S. (2019). Probing the early phase of rapid instructed rule encoding. *Biosystems*, *184*, 103993. https://doi.org/10.1016/j.biosystems.2019.103993

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Canty, A. S., Ripley, B. R., & Brazzale, A. R. (2024). *boot: Bootstrap Functions (Originally by Angelo Canty for S)* (1.3-30) [Computer software]. https://cran.r-project.org/web/packages/boot/index.html

Cole, M. W., Braver, T. S., & Meiran, N. (2017). The task novelty paradox: Flexible control of inflexible neural pathways during rapid instructed task learning. *Neuroscience & Biobehavioral Reviews*, *81*, 4–15. https://doi.org/10.1016/j.neubiorev.2017.02.009

Cole, M. W., Laurent, P., & Stocco, A. (2013). Rapid instructed task learning: A new window into the human brain's unique capacity for flexible cognitive control. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(1), 1–22. https://doi.org/10.3758/s13415-012-0125-7

Corneille, O., & Béna, J. (2023). Instruction-based Replication Studies Raise Challenging Questions for Psychological Science. *Collabra: Psychology*, *9*(1), 82234. https://doi.org/10.1525/collabra.82234

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163–191. https://doi.org/10.1037/0033-2909.104.2.163

Drewe, E. A. (1974). The Effect of Type and Area of Brain Lesion on Wisconsin Card Sorting Test Performance. *Cortex*, *10*(2), 159–170. https://doi.org/10.1016/S0010-9452(74)80006-7

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the Frontal Lobe: The Organization of Goal-Directed Behavior. *Cognitive Psychology*, *30*(3), 257–303. https://doi.org/10.1006/cogp.1996.0008

Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., Bishop, S., & Nimmo-Smith, I. (2008). Goal neglect and Spearman's g: Competing parts of a complex task. *Journal of Experimental Psychology: General*, *137*(1), 131–148. https://doi.org/10.1037/0096-3445.137.1.131

Gabry, J., Češnovar, R., & Johnson, A. (2023). *cmdstanr: R Interface to 'CmdStan'* [Computer software]. https://mc-stan.org/cmdstanr/index.html

Gabry, J., Mahr, T., Bürkner, P.-C., Modrák, M., Barrett, M., Weber, F., Sroka, E. C., Sailynoja, T., & Vehtari, A. (2024). *bayesplot: Plotting for Bayesian Models* (1.11.1) [Computer software]. https://cloud.r-project.org/web/packages/bayesplot/index.html

González-García, C., Formica, S., Liefooghe, B., & Brass, M. (2020). Attentional prioritization reconfigures novel instructions into action-oriented task sets. *Cognition*, *194*, 104059. https://doi.org/10.1016/j.cognition.2019.104059

Hülür, G., Keye-Ehing, D., Oberauer, K., & Wilhelm, O. (2019). *The Effect of Stimulus-Response Compatibility on the Association of Fluid Intelligence and Working Memory with Choice Reaction Times* (1). *2*(1), Article 1. https://doi.org/10.5334/joc.66

Iannone, R., Cheng, J., Schloerke, B., Hughes, E., Lauer, A., & Seo, J. (2023). *gt: Easily create presentation-ready display tables* [Computer software]. https://CRAN.R-project.org/package=gt

Kassambara, A. (2023). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* (0.7.2) [Computer software]. https://cran.r-project.org/web/packages/rstatix/index.html

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin*, *136*(5), 849–874. https://doi.org/10.1037/a0019842

Koch, I. (2001). Automatic and intentional activation of task sets. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1474–1486. https://doi.org/10.1037/0278-7393.27.6.1474

Koch, I., & Kiesel, A. (2022). Task Switching: Cognitive Control in Sequential Multitasking. In A. Kiesel, L. Johannsen, I. Koch, & H. Müller (Eds.), *Handbook of Human Multitasking* (pp. 85–143). Springer International Publishing. https://doi.org/10.1007/978-3-031-04760-2_3

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility--A model and taxonomy. *Psychological Review*, *97*(2), 253. https://doi.org/10.1037/0033-295X.97.2.253

Leeuw, J. R. de, Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an Open-Source Collaborative Ecosystem of Behavioral Experiments. *Journal of Open Source Software*, *8*(85), 5351. https://doi.org/10.21105/joss.05351

Lenth, R. V., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.10.1) [Computer software]. https://cran.r-project.org/web/packages/emmeans/index.html

Lewis-Peacock, J. A., Kessler, Y., & Oberauer, K. (2018). The removal of information from working memory. *Annals of the New York Academy of Sciences*, *1424*(1), 33–44. https://doi.org/10.1111/nyas.13714

Liefooghe, B., & De Houwer, J. (2018). Automatic effects of instructions do not require the intention to execute these instructions. *Journal of Cognitive Psychology*, *30*(1), 108–121. https://doi.org/10.1080/20445911.2017.1365871

Liefooghe, B., De Houwer, J., & Wenke, D. (2013). Instruction-based response activation depends on task preparation. *Psychonomic Bulletin & Review*, *20*(3), 481–487. https://doi.org/10.3758/s13423-013-0374-7

Liefooghe, B., & Verbruggen, F. (2019). On the Assimilation of Instructions: Stimulus-response Associations are Implemented but not Stimulus-task Associations. *Journal of Cognition*, *2*(1), 20. https://doi.org/10.5334/joc.78

Liefooghe, B., Wenke, D., & De Houwer, J. (2012). Instruction-based task-rule congruency effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1325. https://doi.org/10.1037/a0028148

Liesefeld, H. R., & Janczyk, M. (2019). Combining speed and accuracy to control for speed-accuracy trade-offs(?). *Behavior Research Methods*, *51*(1), 40–60. https://doi.org/10.3758/s13428-018-1076-x

Liesefeld, H. R., & Janczyk, M. (2023). Same same but different: Subtle but consequential differences between two measures to linearly integrate speed and accuracy (LISAS vs.

BIS). *Behavior Research Methods*, *55*(3), 1175–1192. https://doi.org/10.3758/s13428-022-01843-2

Lin, Y., Brough, R. E., Tay, A., Jackson, J. J., & Braver, T. S. (2022). Working memory capacity preferentially enhances implementation of proactive control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001195

Luria, A. R. (1973). Chapter 1—THE FRONTAL LOBES AND THE REGULATION OF BEHAVIOR. In K. H. Pribram & A. R. Luria (Eds.), *Psychophysiology of the Frontal Lobes* (pp. 3–26). Academic Press. https://doi.org/10.1016/B978-0-12-564340-5.50006-8

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

Marcovitch, S., Boseovski, J. J., Knapp, R. J., & Kane, M. J. (2010). Goal Neglect and Working Memory Capacity in 4- to 6-Year-Old Children. *Child Development*, *81*(6), 1687–1695. https://doi.org/10.1111/j.1467-8624.2010.01503.x

Meiran, N. (1996). Reconfiguration of Processing Mode Prior to Task Performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1423–1442. https://doi.org/10.1037/0278-7393.22.6.1423

Meiran, N., Cole, M., & Braver, T. (2012). When planning results in loss of control: Intention-based reflexivity and working-memory. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00104

Meiran, N., Pereg, M., Kessler, Y., Cole, M. W., & Braver, T. S. (2015). The power of instructions: Proactive configuration of stimulus–response translation. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *41*(3), 768–786. https://doi.org/10.1037/xlm0000063

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2024). *BayesFactor: Computation of Bayes Factors for Common Designs* (0.9.12-4.7) [Computer software]. https://cran.r-project.org/web/packages/BayesFactor/index.html

Navarro, D. (2021). *lsr: Companion to 'Learning Statistics with R'* (0.5.2) [Computer software]. https://cran.r-project.org/web/packages/lsr/index.html

Nicholson, R., Karayanidis, F., Davies, A., & Michie, P. T. (2006). Components of task-set reconfiguration: Differential effects of '*switch-to*' and '*switch-away*' cues. *Brain Research*, *1121*(1), 160–176. https://doi.org/10.1016/j.brainres.2006.08.101

Oberauer, K. (2009). Chapter 2 Design for a Working Memory. In *Psychology of Learning and Motivation* (Vol. 51, pp. 45–100). Academic Press. https://doi.org/10.1016/S0079-7421(09)51002-X

Ooms, J., Lang, D. T., & Hilaiel, L. (2023). *jsonlite: A Simple and Robust JSON Parser and Generator for R* (1.8.8) [Computer software]. https://cran.r-project.org/web/packages/jsonlite/index.html

Pedersen, T. L. (2024). *patchwork: The composer of plots* (1.2.0) [Computer software]. https://CRAN.R-project.org/package=patchwork

Pereg, M., & Meiran, N. (2019). Rapid instructed task learning (but not automatic effects of instructions) is influenced by working memory load. *PLOS ONE*, *14*(6), e0217681. https://doi.org/10.1371/journal.pone.0217681

R Core Team. (2024). *R: The R Project for Statistical Computing* (4.3.3) [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Roberts, G., & Anderson, M. (2014). Task structure complexity and goal neglect in typically developing children. *Journal of Experimental Child Psychology*, *120*, 59–72. https://doi.org/10.1016/j.jecp.2013.11.003

Robinson, D., Hayes, A., Couch, S., Posit Software PBC, Patil, I., Chiu, D., Gomez, M., Demeshev, B., Menne, D., Nutter, B., Johnston, L., Bolker, B., Briatte, F., Arnold, J., Gabry, J., Selzer, L., Simpson, G., Preussner, J., Hesselberth, J., … Reinhart, A. (2023). *broom: Convert Statistical Objects into Tidy Tibbles* (1.0.5) [Computer software]. https://cran.r-project.org/web/packages/broom/index.html

Rogers, R. D., & Monsell, S. (1995). Costs of a Predictable Switch Between Simple Cognitive Tasks. *Journal of Expferimental Psychology: General*, *124*(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207

Ruge, H., Karcz, T., Mark, T., Martin, V., Zwosta, K., & Wolfensteller, U. (2018). On the efficiency of instruction-based rule encoding. *Acta Psychologica*, *184*, 4–19. https://doi.org/10.1016/j.actpsy.2017.04.005

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do “Brain-Training” Programs Work? *Psychological Science in the Public Interest*, *17*(3), 103–186. https://doi.org/10.1177/1529100616661983

Singmann, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M. S., Højsgaard, S., Fox, J., Lawrence, M. A., Mertens, U., Love, J., Lenth, R., & Christensen, R. H. B. (2024). *afex: Analysis of Factorial Experiments* (1.3-1) [Computer software]. https://cran.r-project.org/web/packages/afex/index.html

Solymos, P., Zawadzki, Z., Bengtsson, H., & R Core Team. (2023). *pbapply: Adding Progress Bar to ‘*apply’ Functions* (1.7-2) [Computer software]. https://cran.r-project.org/web/packages/pbapply/index.html

Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception, & Psychophysics*, *78*(7), 1839–1860. https://doi.org/10.3758/s13414-016-1108-5

Stan Development Team. (2023). *Stan Modeling Language Users Guide and Reference Manual* [Computer software]. //mc-stan.org/

Stuss, D. T., Levine, B., Alexander, M. P., Hong, J., Palumbo, C., Hamer, L., Murphy, K. J., & Izukawa, D. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: Effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia*, *38*(4), 388–402. https://doi.org/10.1016/S0028-3932(99)00093-7

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A New and Improved Word Frequency Database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

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

Verbruggen, F., Liefooghe, B., Vandierendonck, A., & Demanet, J. (2007). Short cue presentations encourage advance task preparation: A recipe to diminish the residual switch cost. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 342–356. https://doi.org/10.1037/0278-7393.33.2.342

Wendt, M., Luna-Rodriguez, A., Reisenauer, R., Jacobsen, T., & Dreisbach, G. (2012). Sequential Modulation of Cue Use in the Task Switching Paradigm. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00287

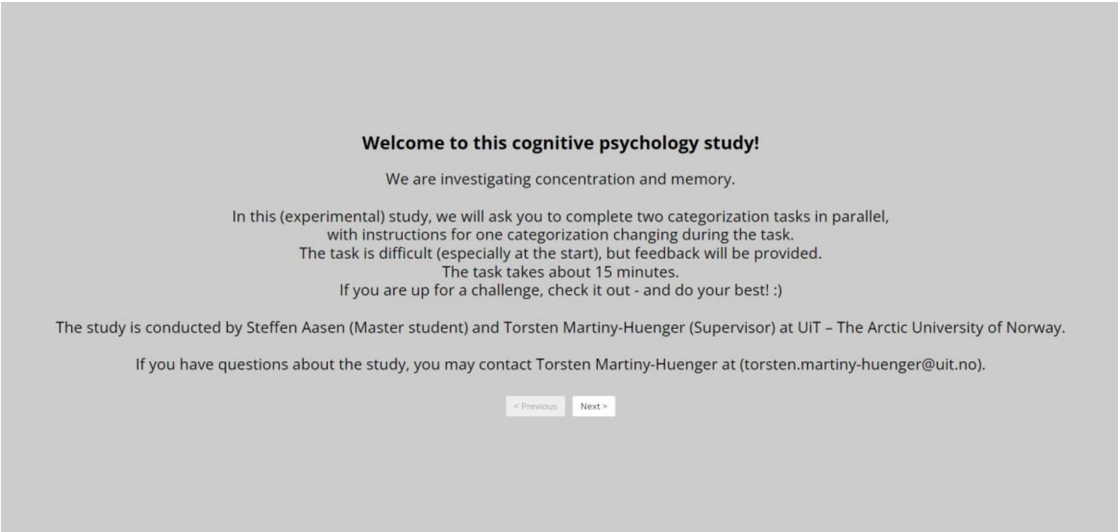
Whitehead, P. S., & Egner, T. (2018). Frequency of prospective use modulates instructed task-set interference. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(12), 1970–1980. https://doi.org/10.1037/xhp0000586

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

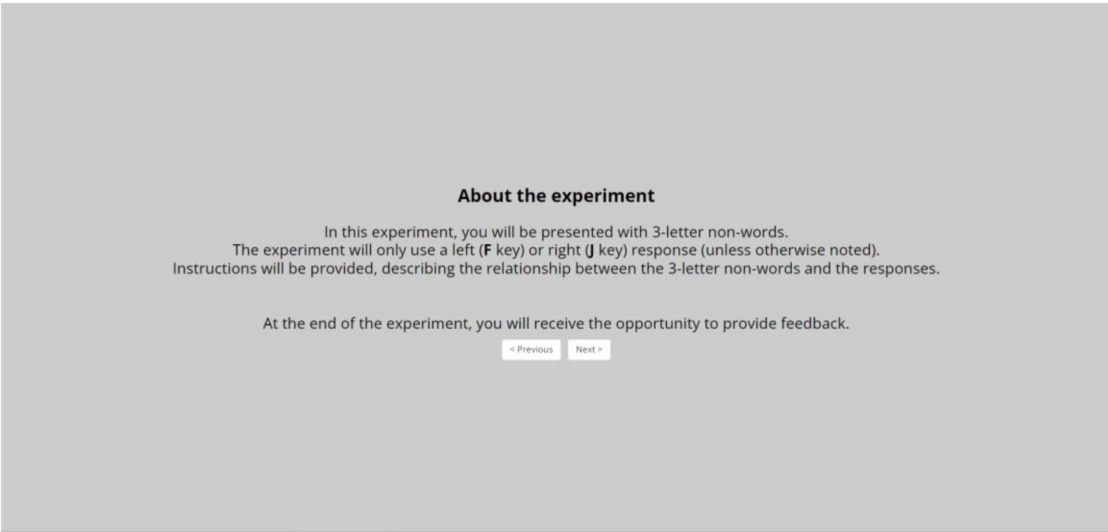
Appendix A

Experiment 1 Instructions and Screens

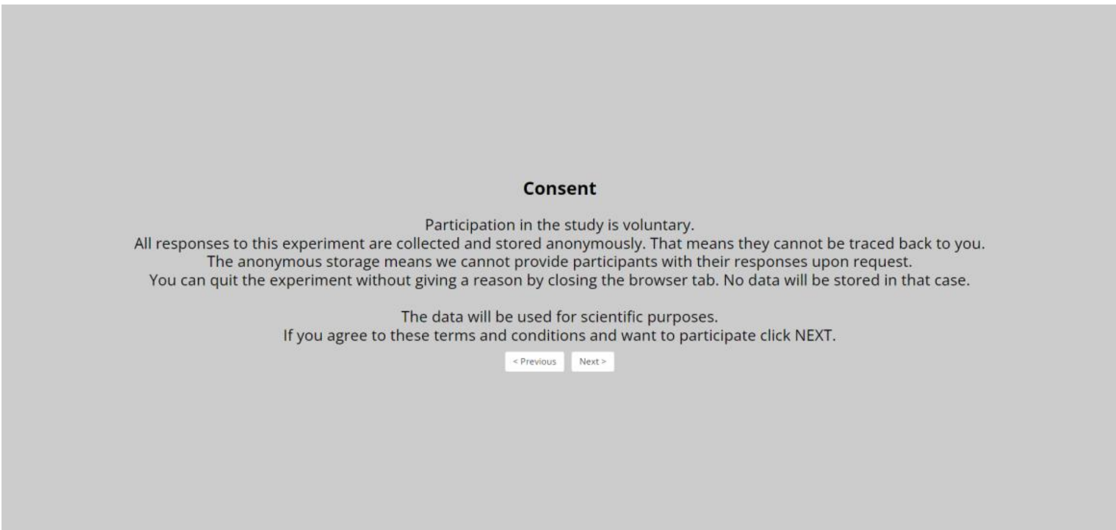
a)



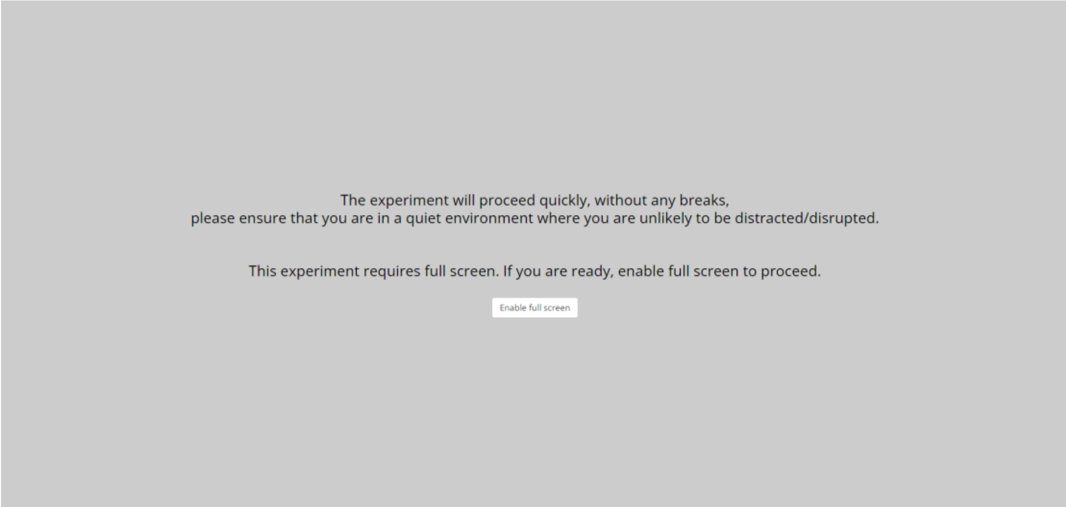
b)



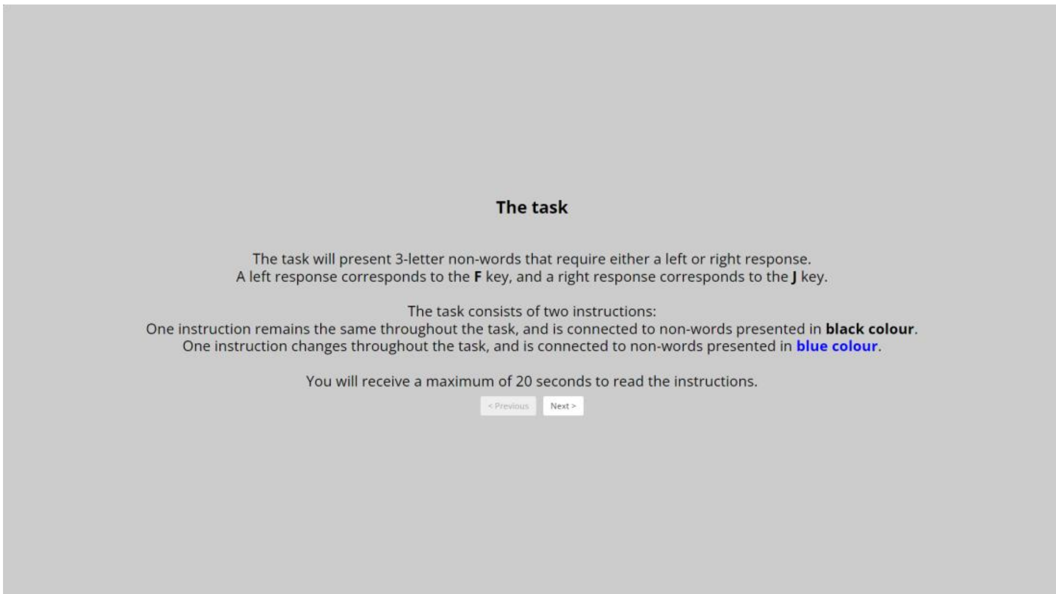
c)



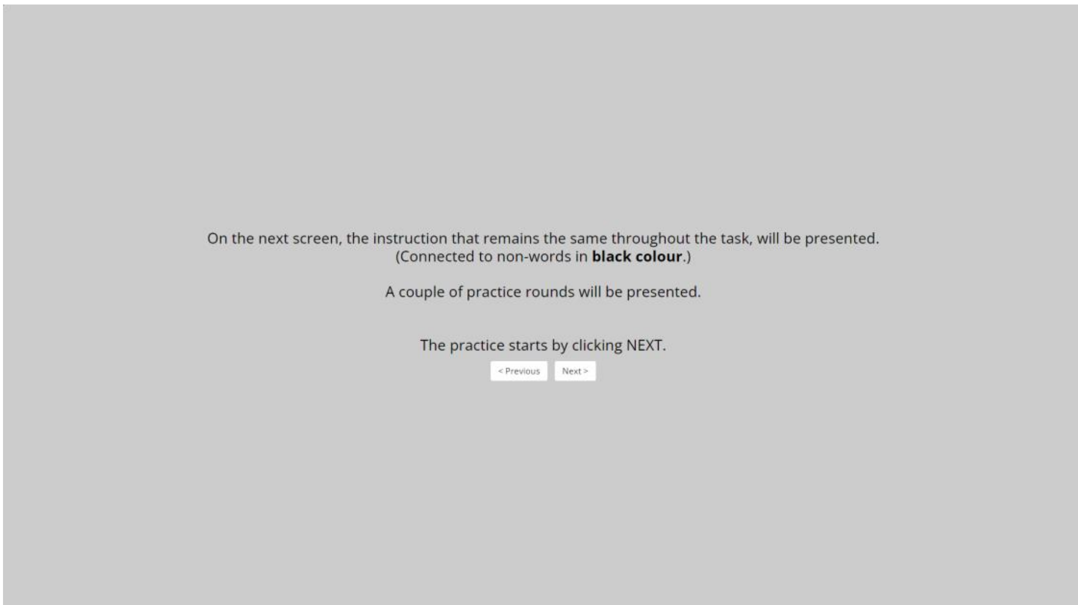
d)


The experiment will proceed quickly, without any breaks,
please ensure that you are in a quiet environment where you are unlikely to be distracted/disrupted.
This experiment requires full screen. If you are ready, enable full screen to proceed.
Enable full screen

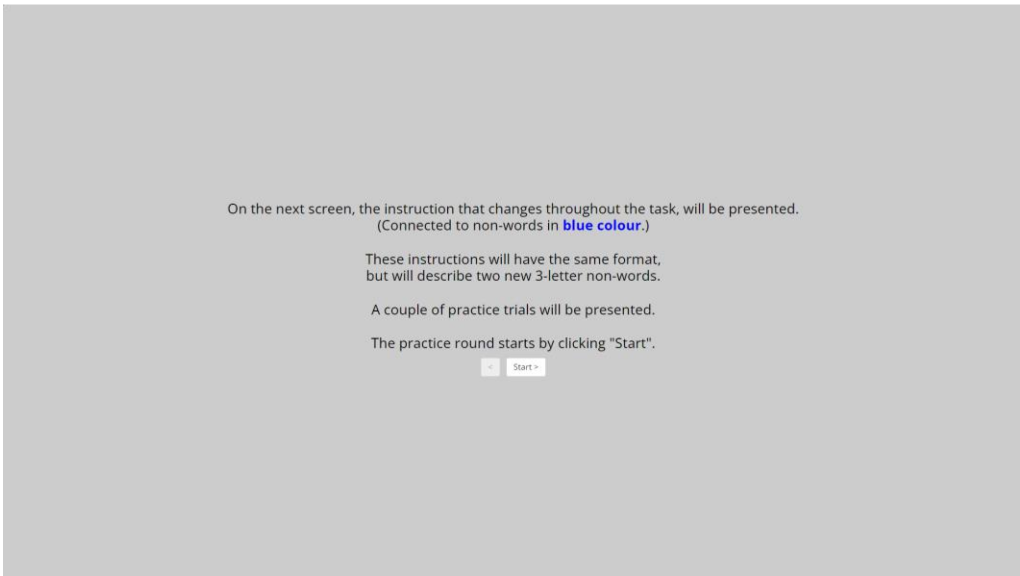
e)


The task
The task will present 3-letter non-words that require either a left or right response.
A left response corresponds to the F key, and a right response corresponds to the J key.
The task consists of two instructions:
One instruction remains the same throughout the task, and is connected to non-words presented in black colour.
One instruction changes throughout the task, and is connected to non-words presented in blue colour.
You will receive a maximum of 20 seconds to read the instructions.
< Previous
Next >

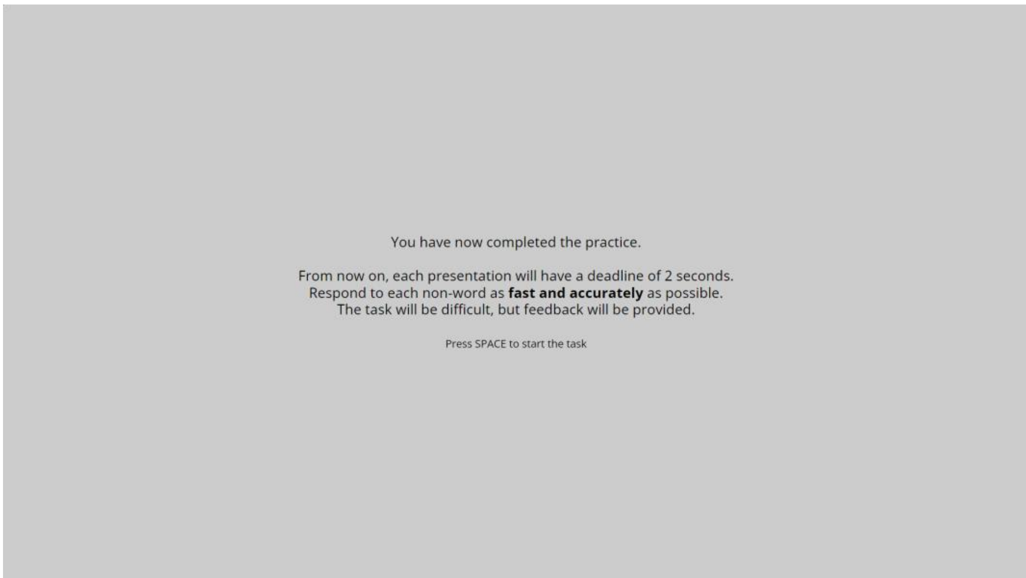
f)


On the next screen, the instruction that remains the same throughout the task, will be presented.
(Connected to non-words in black colour.)
A couple of practice rounds will be presented.
The practice starts by clicking NEXT.
< Previous
Next >

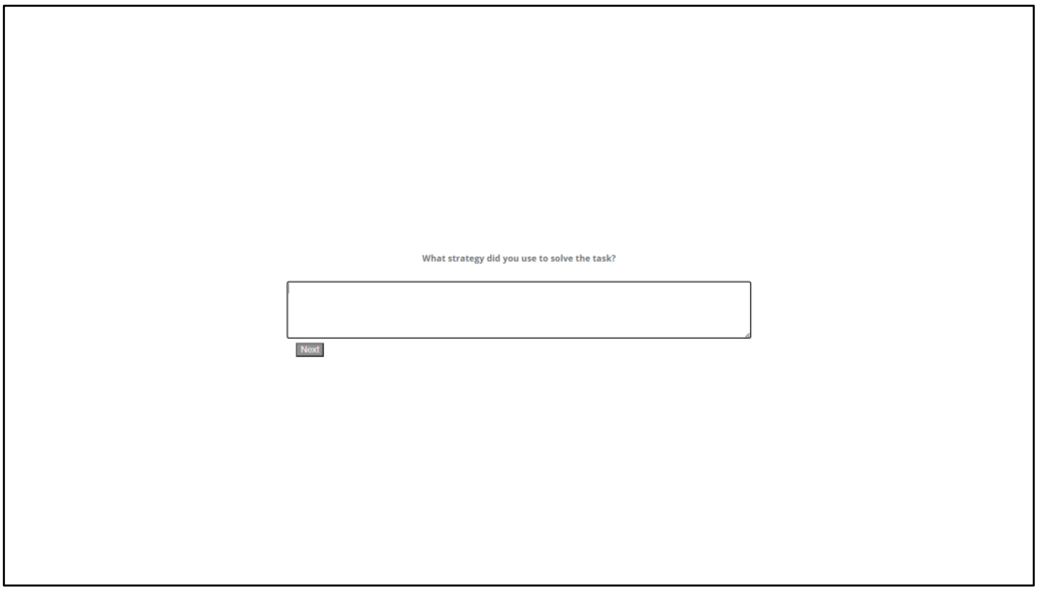
g)


On the next screen, the instruction that changes throughout the task, will be presented.
(Connected to non-words in **blue colour**.)
These instructions will have the same format,
but will describe two new 3-letter non-words.
A couple of practice trials will be presented.
The practice round starts by clicking "Start".
<
Start >

h)


You have now completed the practice.
From now on, each presentation will have a deadline of 2 seconds.
Respond to each non-word as **fast and accurately** as possible.
The task will be difficult, but feedback will be provided.
Press SPACE to start the task

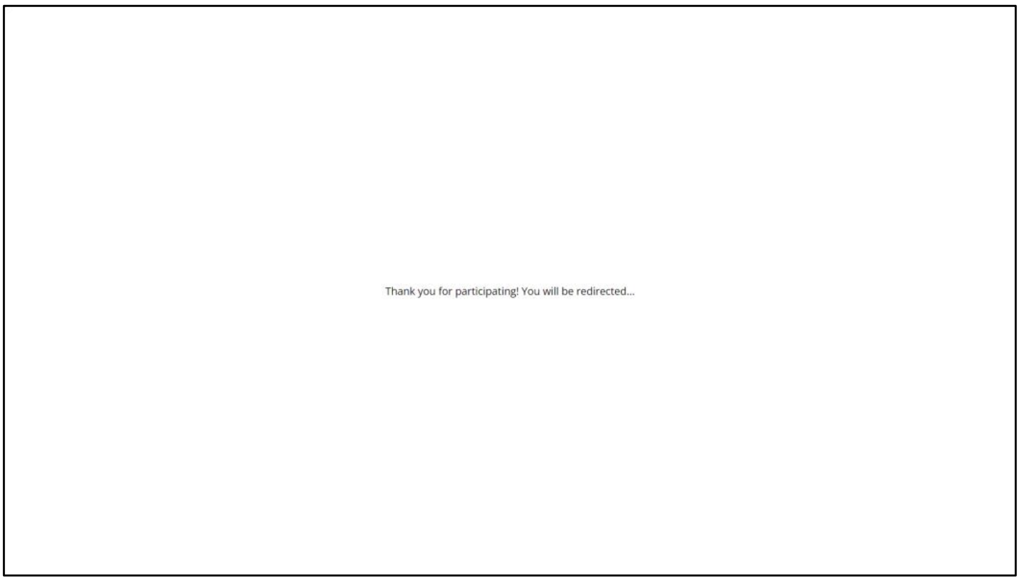
i)


What strategy did you use to solve the task?
Next

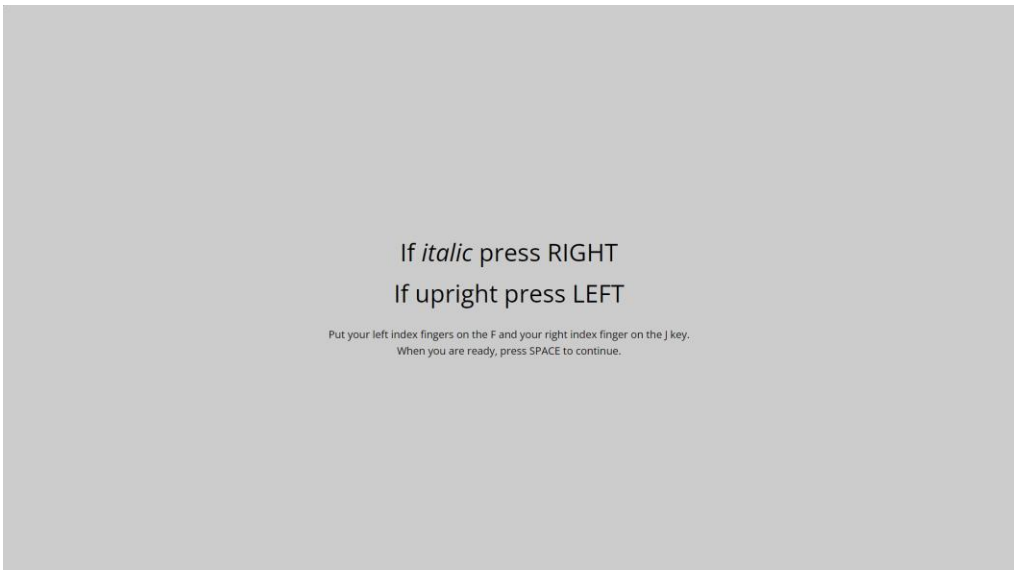
j)


Do you have any comments, thoughts, or remarks in relation to the experiment?
Back
End experiment

k)


Thank you for participating! You will be redirected...

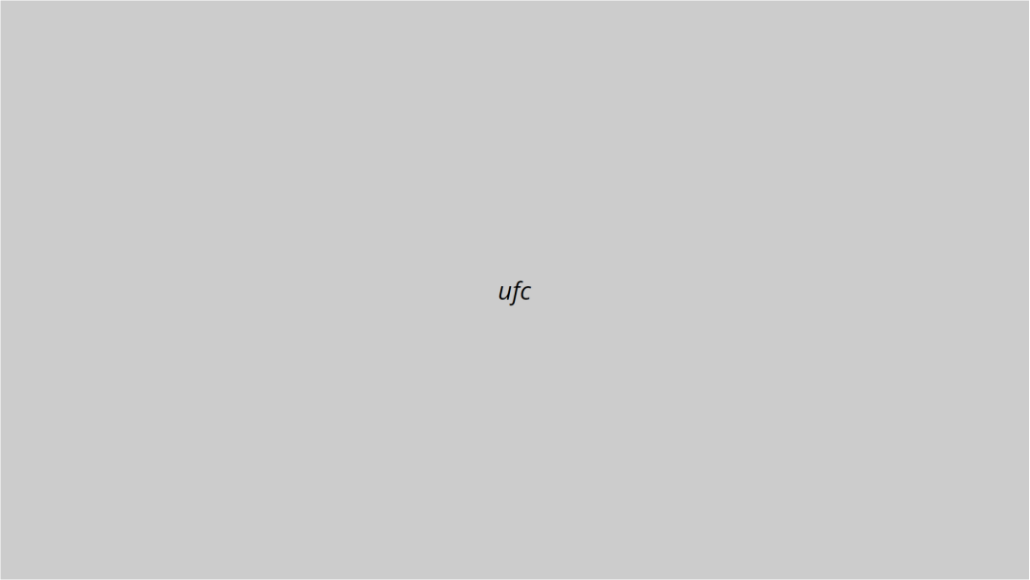
l)


If *italic* press RIGHT
If upright press LEFT
Put your left index fingers on the F and your right index finger on the J key.
When you are ready, press SPACE to continue.

m)


If ufc press LEFT
If frn press RIGHT
Put your left index fingers on the F and your right index finger on the J key.
When you are ready, press SPACE to continue.

n)

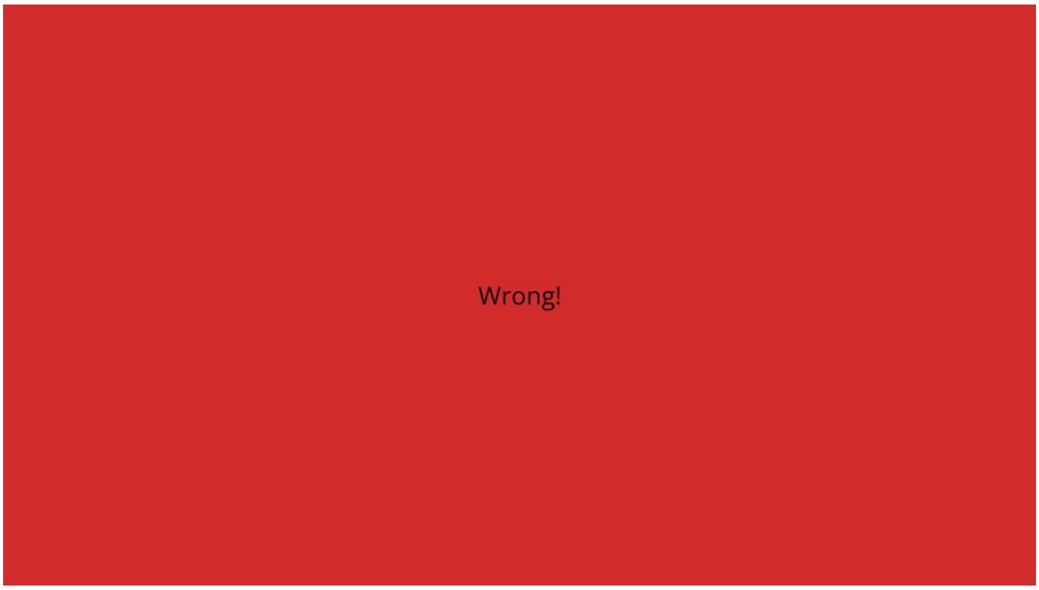

ufc

o)

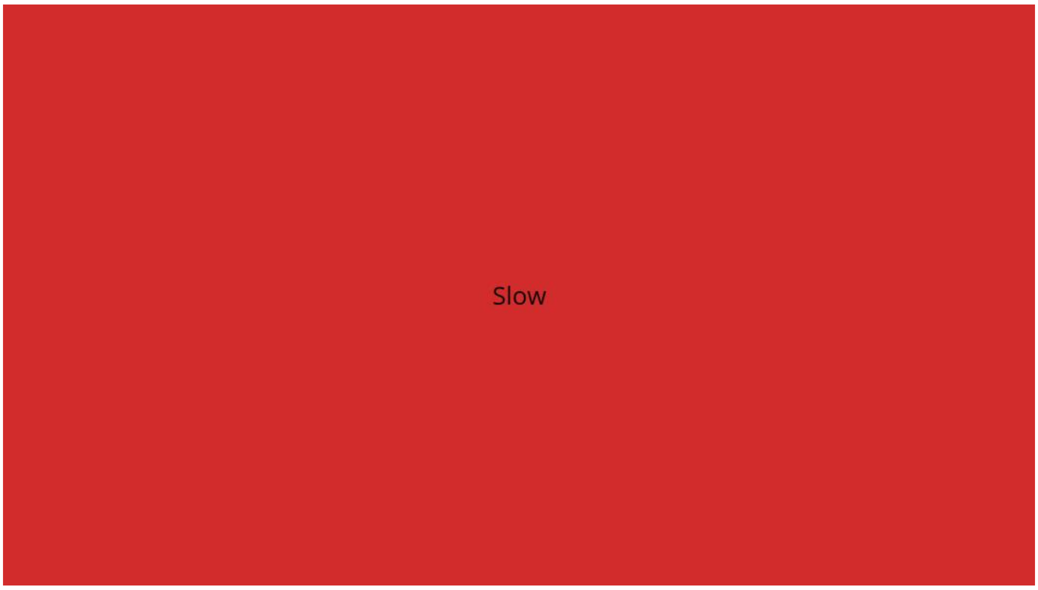

ufc

p)

q)



r)



Note. (a-h) General information, instructions, and consent screens. (i-j) Open feedback screens (a black outline has been added to illustrate the screen). (k) Redirect screen. Example of the (l) diagnostic instruction, (m) inducer instruction, (n) diagnostic italic trial, (o) diagnostic upright trial and the (p) inducer trial. Feedback screens relating to a (q) wrong amd (r) slow response.

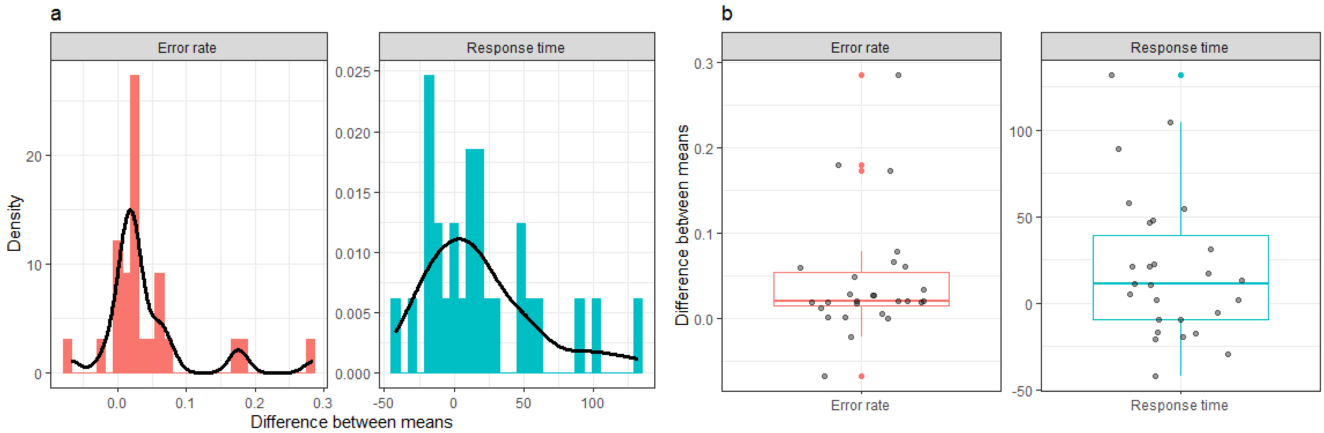
Appendix B

Robustness Checks of the Experiment 1 Results

We include two non-parametric tests of the difference between the means due some outliers in the difference in means for response time and errors (Figure B1). To do this, we did one paired sample Wilcoxon signed ranked test for the outcome response time and error rate. Results indicated a significant difference in the median for response time (V = 276, p = .018), and for error rate (V = 316, p < .001).

Figure B1

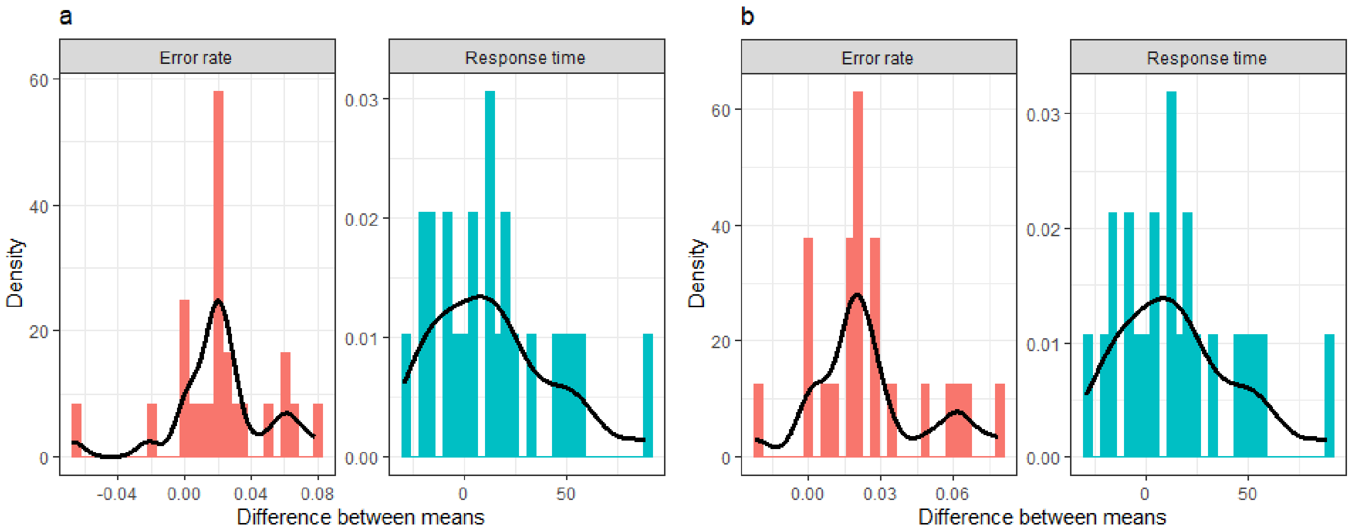
(a) Histogram and (b) Boxplot of the Difference Between the Means for Response Time and Error Rate



To investigate whether the extreme outliers had any effect on the results, we did four additional paired sample t-tests. Two of the tests related to testing the data after removing the extreme outliers, and two after removing all outliers (see Figure B2 for the distribution after removing the outliers). Removing the extreme outliers still resulted in a significant difference between the means for response time (t(23) = 2.25, p = .017), and error rate (t(23) = 3.47, p = .001). Removing the outliers still resulted in a significant difference between the means for response time (t(22) = 2.46, p = .011), and error rate (t(22) = 5.07, p < .001). These results indicate that, regardless of the influence of the outliers, the results still remain significant, and our results are therefore relatively robust.

Figure B2

Distribution of the Difference in Means After Excluding the (a) Extreme Outliers and (b) all Outliers



In addition, I did two bootstraps repeated 5000 times of the difference between the means for response times and error rate. For response time, the bootstrap revealed a 95 % confidence interval [3.41, 34.7] excluding 0. Similarly, for error rate, the bootstrap revealed a 95 % confidence interval [0.02, 0.07] also excluding 0. These results suggest that the difference between means for both response time and error rate is likely to be different from 0.

Appendix C

Test Statistics for Response Time (RT), Error Rate (PE) and the Integrated Scores

	Incongruent		Congruent					Bayesian			
Variable	*M*	*SD*	*M*	*SD*	M_{diff}	$t(26)$	*p*	*b*	BF_{10}	HDI	Cohen's *d*
RT	661	111	642	101	19.2	2.41	.012	17.5	2.31	[1.28, 32.9]	0.46
PE	0.08	0.07	0.04	0.03	0.04	3.17	.002	0.04	10.4	[0.01, 0.07]	0.61
LISAS	718	149	668	112	50.7	3.43	.001	46.9	18.2	[17.9, 76.7]	0.66
BIS_{LISAS}	687	151	629	111	57.6	3.50	< .001	53.7	21.5	[19.5, 86.3]	0.67
BIS	-0.23	0.77	0.07	0.57	-0.29	-3.50	< .001	-0.27	21.5	[-0.43, -0.1]	0.67

Note. The integrated scores are the linear integrated speed-accuracy score (LISAS) and the balanced integrated score (BIS). The BIS_{LISAS} is the BIS transformed to response time (Liesefeld & Janczyk, 2023) for direct comparison to the LISAS. M_{diff} = Difference between the means; *b* = estimated Bayesian difference between the means; BF_{10} = Bayesian factor for the alternative hypothesis; HDI = highest density interval.

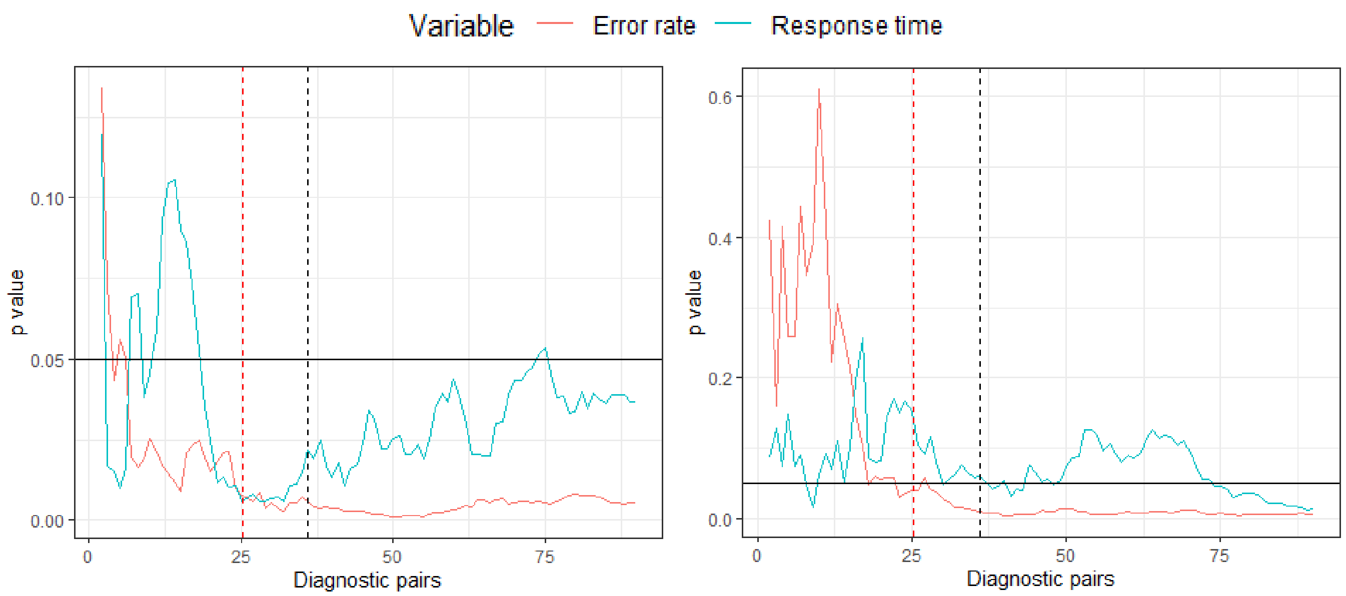
Appendix D

Estimating Power and Post-cue Diagnostic Trials for Experiment 2

To ensure I have enough power to detect an effect in the post-cue run, I resample from the first experiment. After the exclusion criteria, I was left with 168 trials. This corresponds to a loss of 30 %. In other words, whatever post-cue run I decide, I need to estimate an expected loss of about 30%. I determined to use around 36 pairs (72 trials) for the post-cue run, corresponding to an average of 3 trials per post-cue run. These runs are displayed as a black dashed line, and the expected remaining trials (after loss) are indicated with a red dashed line. Results for response time (RT) and proportion of errors (PE) are displayed.

Figure C1

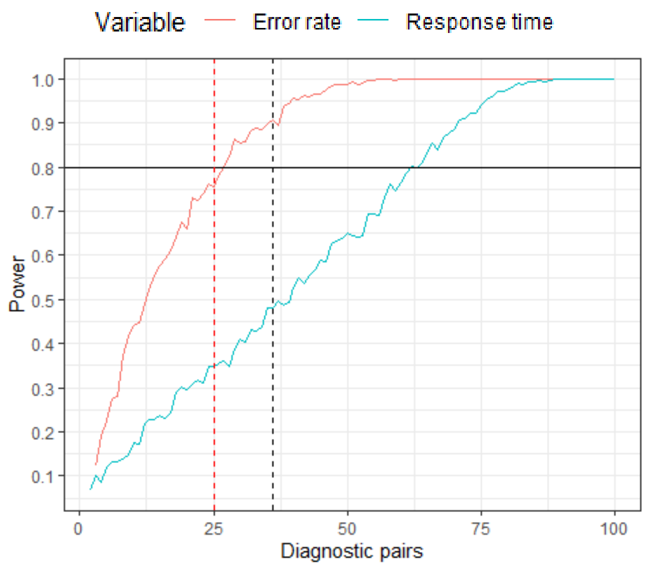
Change in p-Value With Increasing Number of Diagnostics Pairs



Note. The left figure relates to increasing diagnostic pairs as sampled from the start of experiment 1, and the right figure relates to sampling from the end.

Figure C2

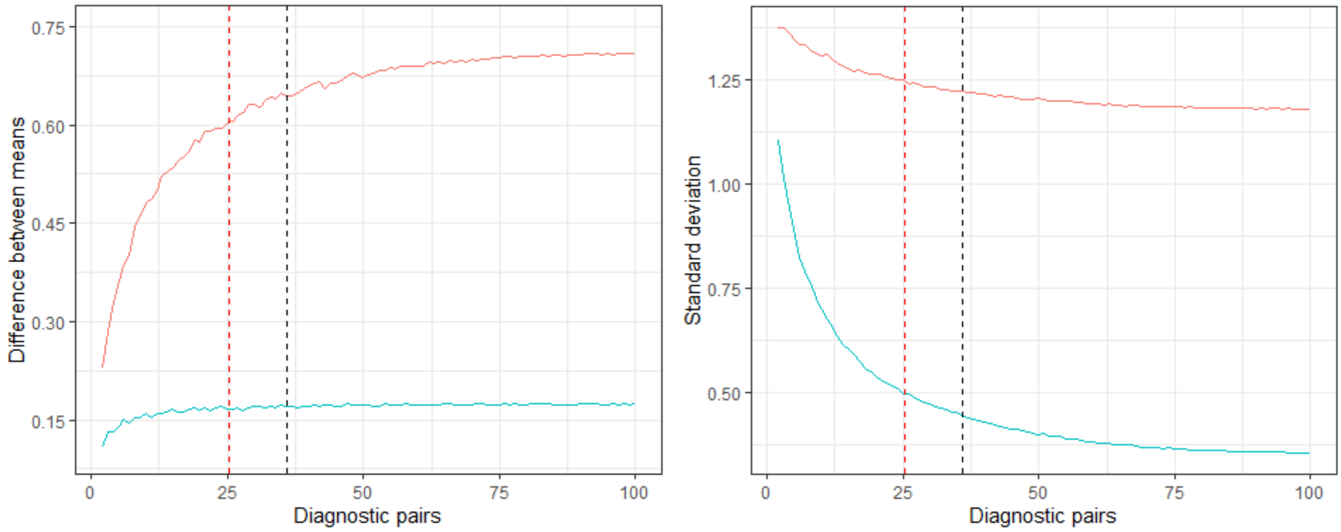
Changes in p-Value With Increasing Number of Diagnostic Pairs Through Randomly Sampling From Experiment 1



Note. Data is randomly sampled from experiment 1 each pair is averaged across 900 resampling steps.

Figure C3

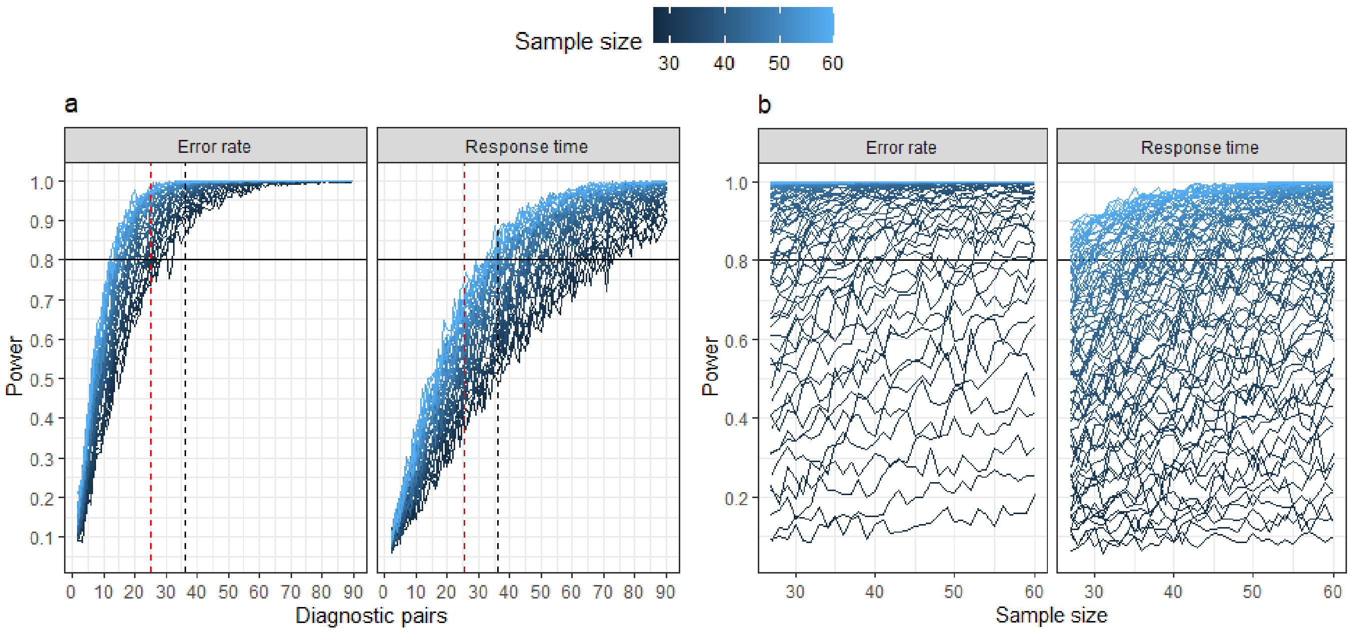
Changes in the Mean Difference and Standard Deviation by Increasing the Number of Diagnostic Pairs



Note. The left figure illustrates the changes in mean difference by increasing number of diagnostic pairs. The right figure illustrates the changes in mean standard deviation difference by increasing the number of diagnostic pairs.

Figure C4

Changes in Power by Increasing the Sample Size and the Number of Diagnostic Pairs



Note. Both parts (a and b) show the same data, but the x and colour are flipped. Each sample size and pair size are resampled 500 times and averaged.

Figure C5

Changes in Power Across Various Sample Sizes Split by the Relevant Diagnostic Pairs

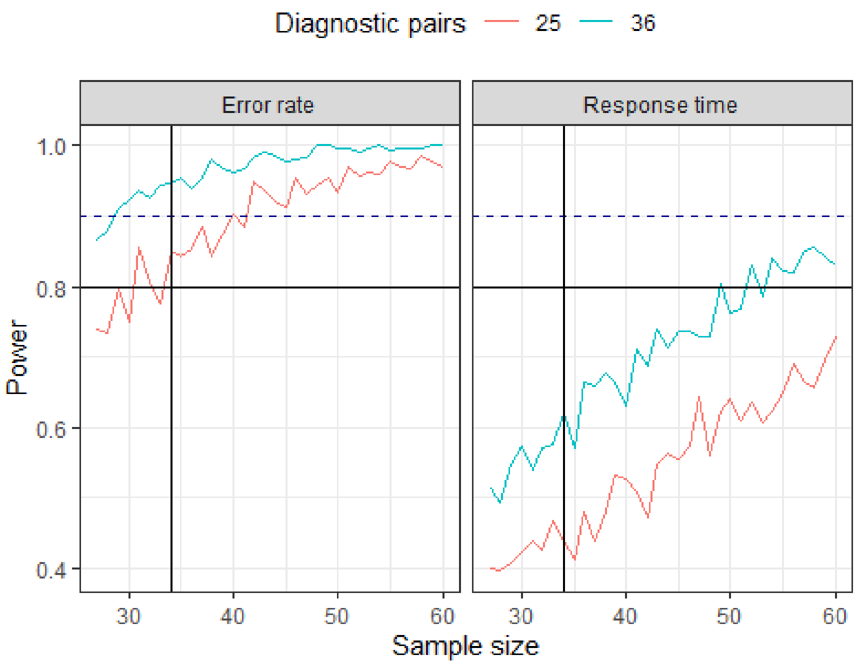
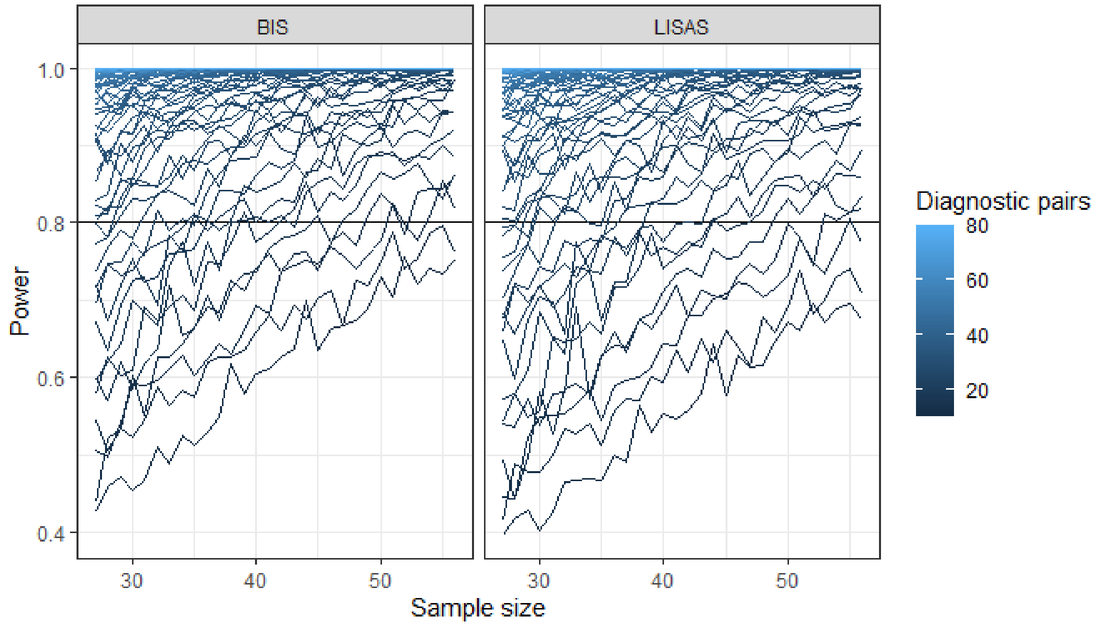


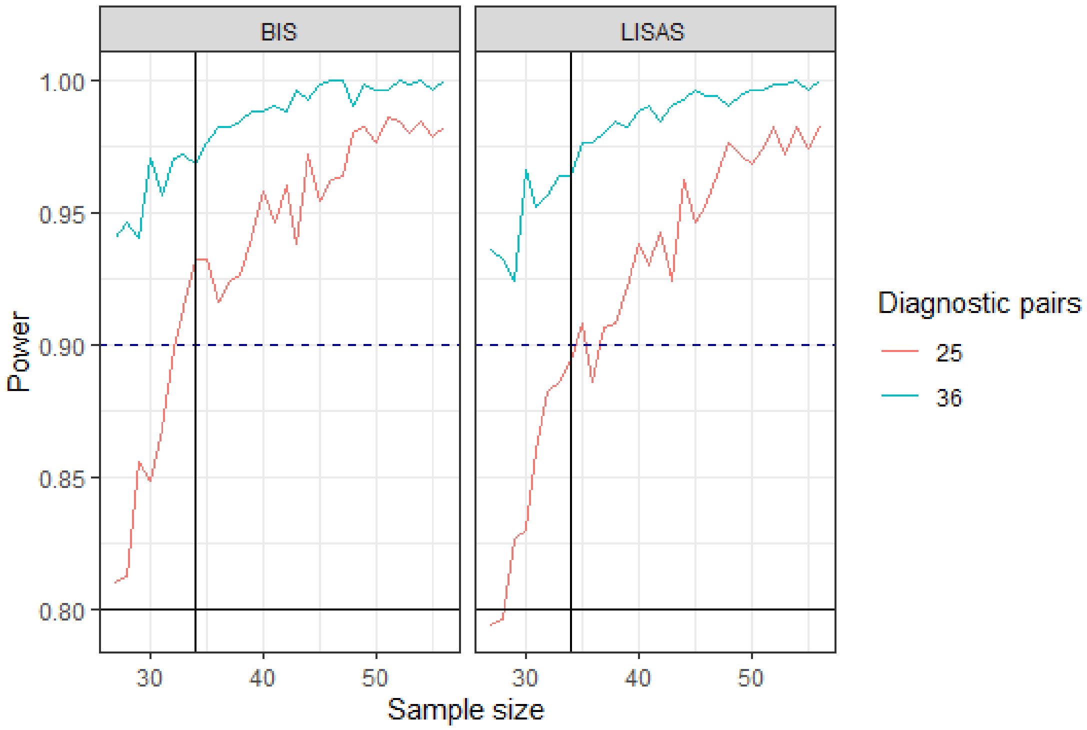
Figure C6

Changes in Power by Increasing the Sample Size for the Integrated Scores

a)



b)



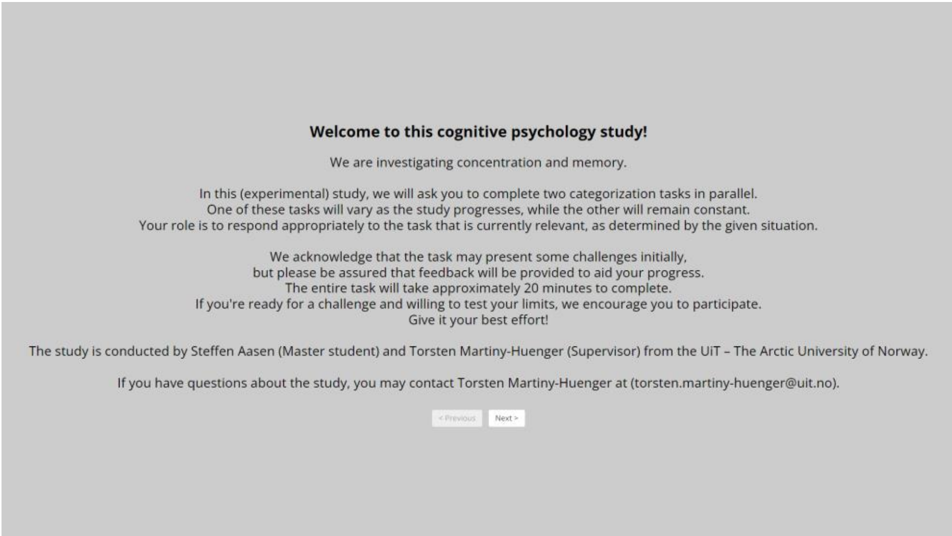
Note. The top graph (a) illustrates the changes in power over increasing sampled sizes by increasing the diagnostic pairs for the linear speed-accuracy integration score (LISAS) and the balanced integration score (BIS). The bottom (b) graph highlights the changes for the diagnostic pairs 36 (72 diagnostic trials) and the pairs after the expected loss (25 pairs, i.e., 50 trials).

Random slicing from the first experiment suggests that I would need slightly more than 30 participants (34) to achieve 80% power for PE. RT, on the other hand, remain below 50% after the expected loss of data. In other words, increasing the sample size would not remedy the problem for RT.

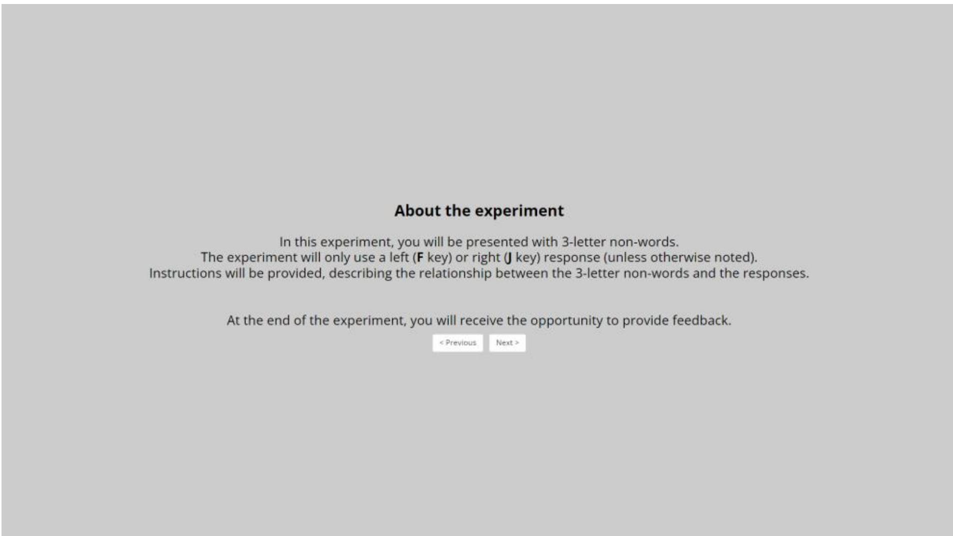
Appendix E

Experiment 2 Instructions and Screens

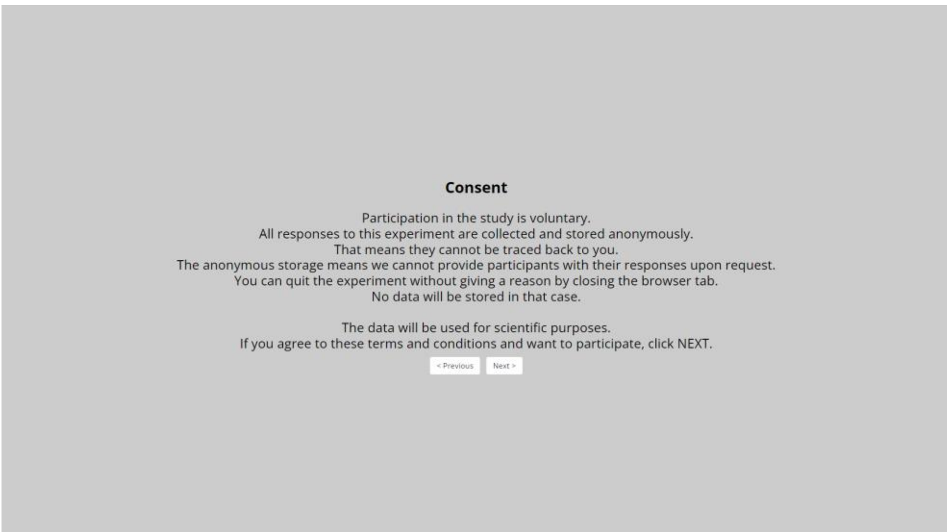
a)



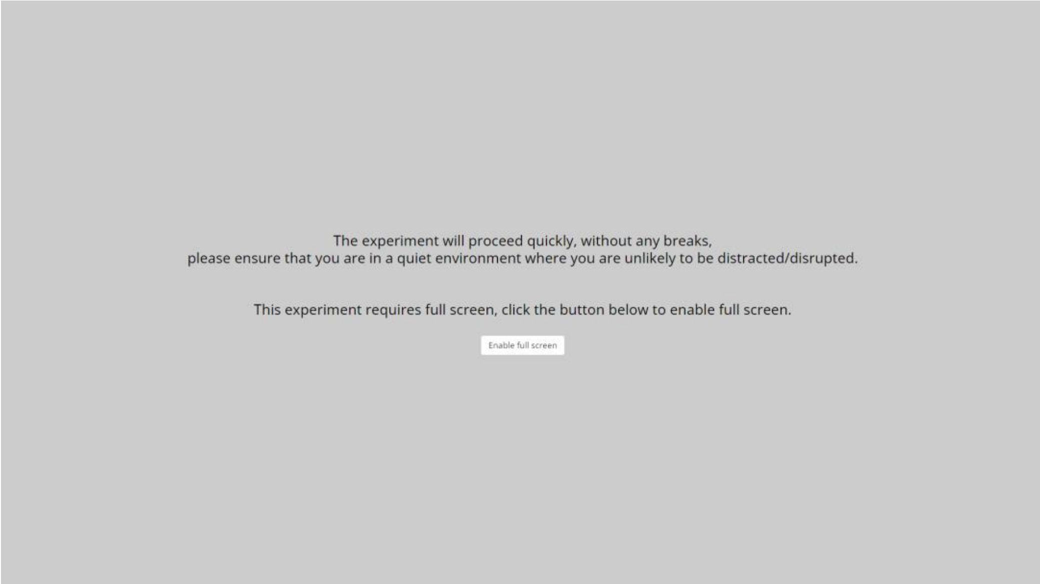
b)



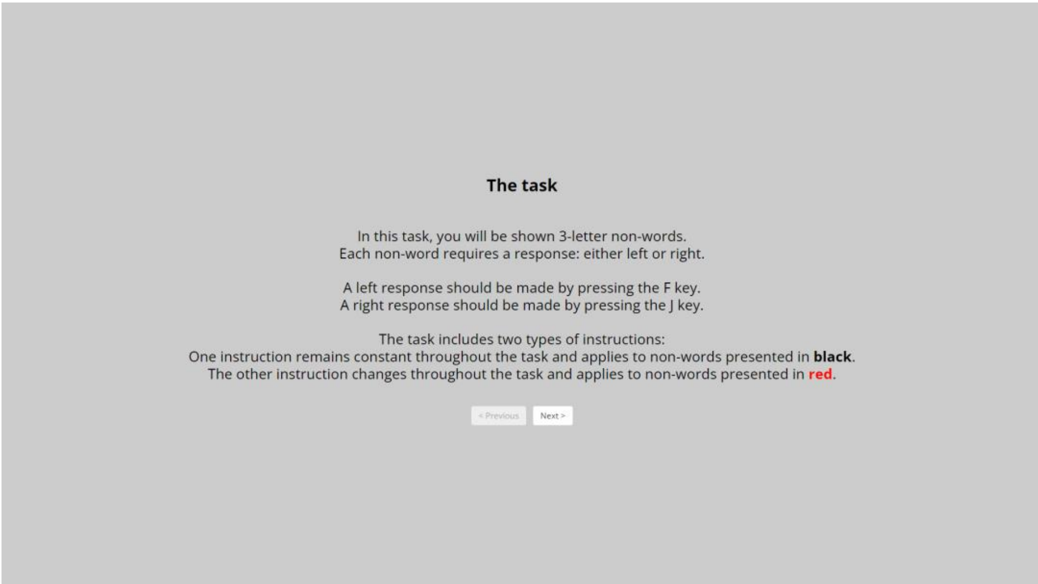
c)



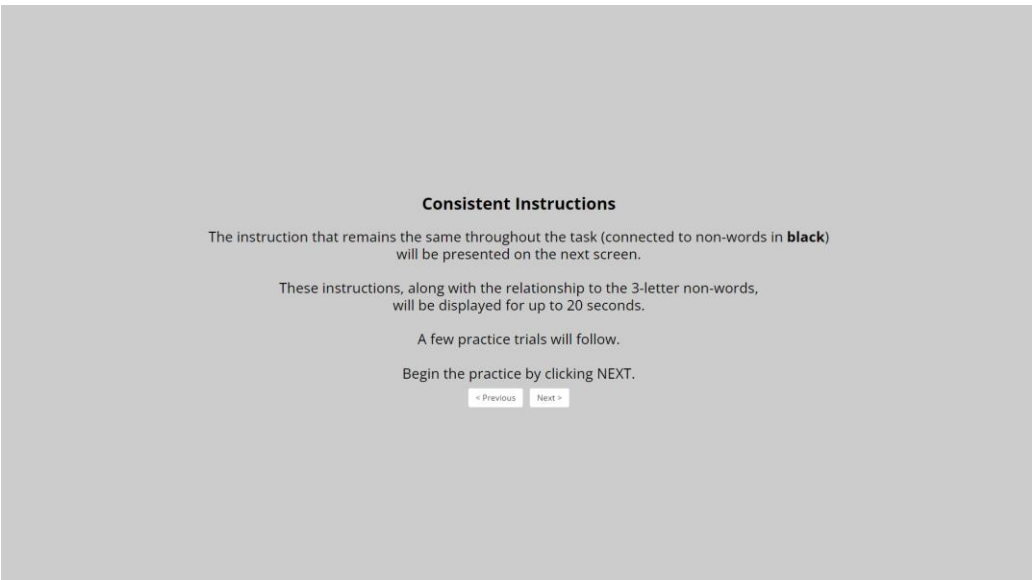
d)


The experiment will proceed quickly, without any breaks,
please ensure that you are in a quiet environment where you are unlikely to be distracted/disrupted.
This experiment requires full screen, click the button below to enable full screen.
Enable full screen

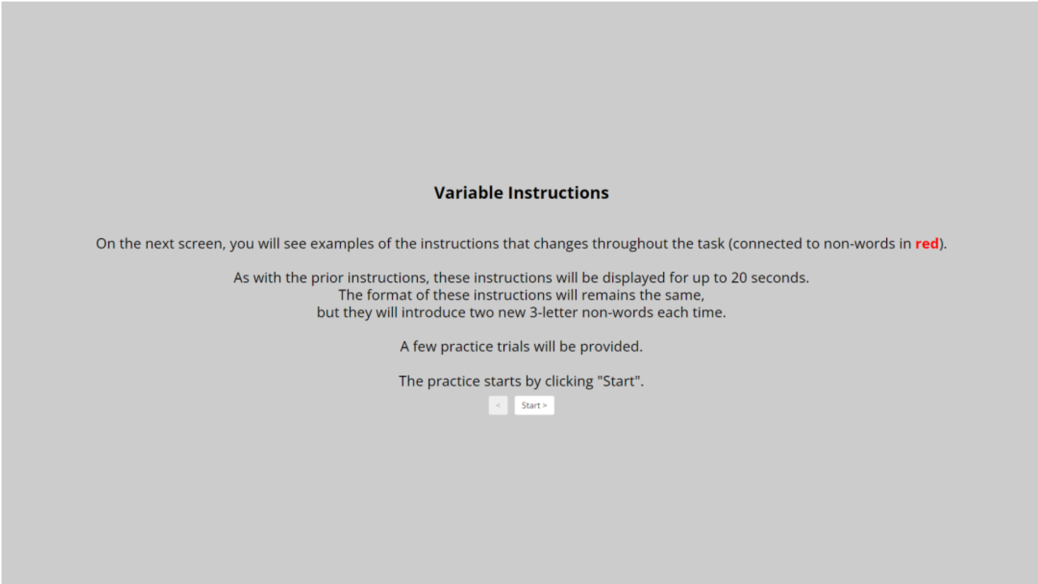
e)


The task
In this task, you will be shown 3-letter non-words.
Each non-word requires a response: either left or right.
A left response should be made by pressing the F key.
A right response should be made by pressing the J key.
The task includes two types of instructions:
One instruction remains constant throughout the task and applies to non-words presented in black.
The other instruction changes throughout the task and applies to non-words presented in red.
< Previous
Next >

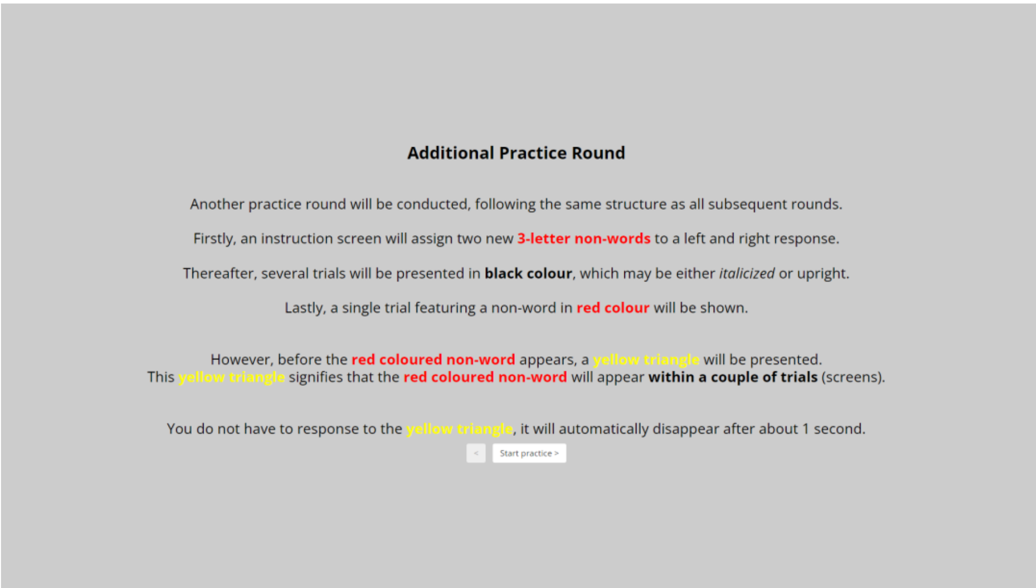
f)


Consistent Instructions
The instruction that remains the same throughout the task (connected to non-words in black)
will be presented on the next screen.
These instructions, along with the relationship to the 3-letter non-words,
will be displayed for up to 20 seconds.
A few practice trials will follow.
Begin the practice by clicking NEXT.
< Previous
Next >

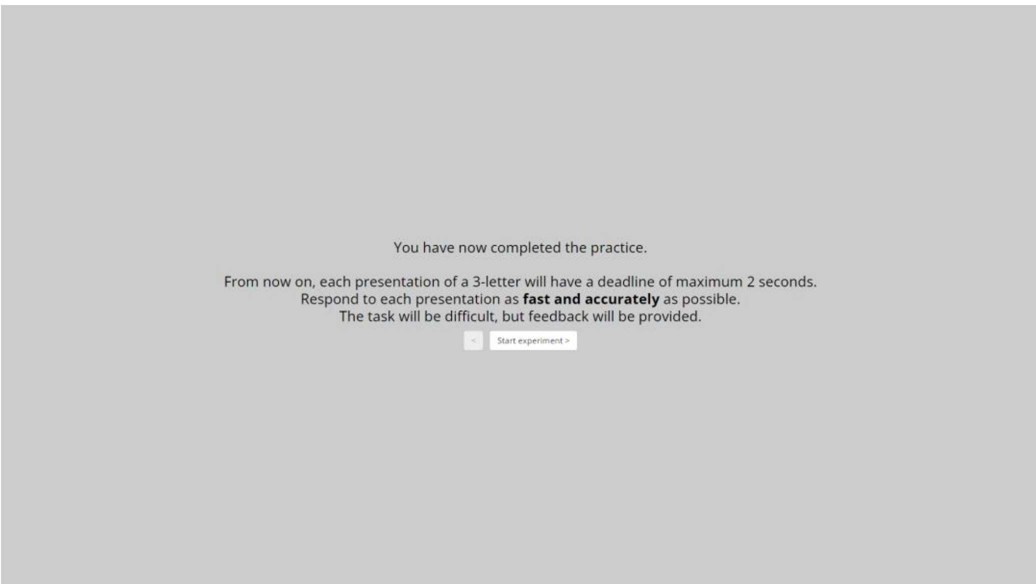
g)


Variable Instructions
On the next screen, you will see examples of the instructions that changes throughout the task (connected to non-words in red).
As with the prior instructions, these instructions will be displayed for up to 20 seconds.
The format of these instructions will remains the same,
but they will introduce two new 3-letter non-words each time.
A few practice trials will be provided.
The practice starts by clicking "Start".
<
Start >

h)


Additional Practice Round
Another practice round will be conducted, following the same structure as all subsequent rounds.
Firstly, an instruction screen will assign two new 3-letter non-words to a left and right response.
Thereafter, several trials will be presented in black colour, which may be either italicized or upright.
Lastly, a single trial featuring a non-word in red colour will be shown.
However, before the red coloured non-word appears, a yellow triangle will be presented.
This yellow triangle signifies that the red coloured non-word will appear within a couple of trials (screens).
You do not have to response to the yellow triangle, it will automatically disappear after about 1 second.
<
Start practice >

i)


You have now completed the practice.
From now on, each presentation of a 3-letter will have a deadline of maximum 2 seconds.
Respond to each presentation as fast and accurately as possible.
The task will be difficult, but feedback will be provided.
<
Start experiment >

j)

Note. (a-d) About the experiment and consent. (e-h) Explanation of the instructions. (i) Practice instructions. (i) Example of the cue. Otherwise, the elements were the same as experiment 1 (Appendix A).

Appendix F

Experiment 2 Tables and Figures

Table F1

Mean (M) and Standard Deviation (SD) for Across cue and Congruency

	Pre-cue				Post-cue			
	Incongruent		Congruent		Congruent		Incongruent	
Dependent variable	M	SD	M	SD	M	SD	M	SD
Response time	653	91.8	634	90.6	688	98.9	662	85.7
Error rate	0.06	0.05	0.04	0.04	0.06	0.06	0.03	0.05
LISAS	701	111	657	94	730	118	688	91.4
BIS	-0.1	0.56	0.14	0.49	-0.25	0.6	-0.01	0.5

Note. LISAS = Linear integrated speed-accuracy score (Vandierendonck, 2017); BIS = Balanced integration score (Liesefeld & Janczyk, 2019).

Table F2

Repeated Measures Analysis of Variance for Each of the Dependent Variable

	Response time			Error rate			LISAS			BIS		
Predictors	F(1,34)	p	η_p^2	F(1,34)	p	η_p^2	F(1,34)	p	η_p^2	F(1,34)	p	η_p^2
Congruency	30.3	< .001	0.471	8.71	.006	0.204	30.8	< .001	0.476	25.2	< .001	0.426
Cue	26.3	< .001	0.437	0.36	.552	0.010	15.1	< .001	0.308	9.67	.004	0.221
Congruency × Cue	0.30	.588	0.009	0.19	.669	0.005	0.01	.942	0.000	0.00	.995	0.000

Note. LISAS = Linear integrated speed-accuracy score (Vandierendonck, 2017); BIS = Balanced integration score (Liesefeld & Janczyk, 2019).

Table F3

Coefficient for the Predictors in the Various Models

	Response time			Error rate			LISAS			BIS		
Predictors	*b*	*t*	*p*	*b*	*t*	*p*	*b*	*t*	*p*	*b*	*t*	*p*
Congruency	19.5	2.64	.047	0.03	2.63	.048	43.5	3.97	< .001	-0.24	-3.63	.002
Cue	-34.5	-4.67	< .001	0.01	0.67	.907	-29.3	-2.68	.042	0.15	2.28	.109
Congruency × Cue	-9.32	-1.26	.589	0.03	2.70	.040	13.0	1.19	.635	-0.09	-1.34	.542

Note. LISAS = Linear integrated speed-accuracy score (Vandierendonck, 2017); BIS = Balanced integration score (Liesefeld & Janczyk, 2019).

Table F4

Minimum and Maximum F, p and Partial-eta Squared for Each of the Dependent Variable and Their Predictors

	$F(1,33)$		p		η_p^2	
Predictors	Min	Max	Max	Min	Min	Max
Response time						
Congruency	27.6	35.9	< .001	< .001	0.46	0.52
Cue	23.9	34	< .001	< .001	0.42	0.51
Congruency × Cue	0.03	0.98	.862	.329	0	0.03
Error rate						
Congruency	7.28	11.8	.011	.002	0.18	0.26
Cue	0.02	0.91	.889	.347	0	0.03
Congruency × Cue	0	0.87	.998	.358	0	0.03
LISAS						
Congruency	28.1	37.8	< .001	< .001	0.46	0.53
Cue	13.3	24.9	< .001	< .001	0.29	0.43
Congruency × Cue	0	0.26	.999	.612	0	0.01
BIS						
Congruency	22.8	32.1	< .001	< .001	0.41	0.49
Cue	8.19	19.1	.007	< .001	0.2	0.37
Congruency × Cue	0	0.21	.999	.648	0	0.01

Note. LISAS = Linear integrated speed-accuracy score (Vandierendonck, 2017); BIS = Balanced integration score (Liesefeld & Janczyk, 2019).

Table F5

Bayesian Models for the Dependent Variables Response Time, Error Rate, LISAS and BIS

	Response Time					Error rate					LISAS					BIS				
				HDI					HDI					HDI					HDI	
Predictors	*b*	p_b	ER_b	Low	High	*b*	p_b	ER_b	Low	High	*b*	p_b	ER_b	Low	High	*b*	p_b	ER_b	Low	High
Congruency	-19.5*	.995	213	-34.1	-4.61	-0.03*	.994	163	-0.05	-0.01	-43.4*	1.00	Inf	-65.4	-21.5	0.24*	>.999	2570	0.11	0.37
Cue	34.5*	1.00	Inf	19.9	49.3	-0.01	.750	3	-0.03	0.01	29.4*	.996	230	7.93	51.4	-0.15*	.988	80.4	-0.28	-0.02
Cue × Congruency	-5.73	.703	2.37	-26.7	14.8	0.01	.667	2	-0.02	0.04	0.91	.522	1.09	-30.4	31.8	0.00	.498	0.99	-0.19	0.19
Model fit																				
Sigma (subject)	31.3*	1.00	Inf	27.1	35.8	0.05*	1.00	Inf	0.04	0.05	46.4*	1.00	Inf	40.2	53.1	0.28*	1.00	Inf	0.24	0.32
R^2	.889			.865	.905	.228			.077	.367	.813			.77	.843	.75			.69	.793
LOOIC	1404			SE = 21.5		-440			SE = 25.5		1514			SE = 24		80.8			SE = 24.2	

Note. p_b = the probability that the effect is in the noted (*b*) direction; ER_b = evidence ratio for the noted (*b*) direction; LISAS = Linear integrated speed-accuracy score (Vandierendonck, 2017); BIS = Balanced integration score (Liesefeld & Janczyk, 2019).

* $p_b > .95$

Table F6

Minimum and Maximum Bayesian Estimation of Sample n = 34

	b		p_b		ER_b		HDI			
							Lower		Upper	
Predictors	Low	High	Low	High	Low	High	Low	High	Low	High
Response time										
Congruency	-21.8*	-17.2*	0.99	>0.99	76.9	719	-36.4	-31.7	-2.43	-2.43
Cue	31.1*	38.3*	>0.99	1.00	8999	Inf	16.7	24.0	52.7	52.7
Congruency × Cue	-9.89	-1.78	0.57	0.83	1.32	4.78	-30.7	-22.1	18.8	18.8
Error rate										
Congruency	-0.03*	-0.03*	0.99	>0.99	92.3	1199	-0.06	-0.05	0.00	0.00
Cue	-0.01	0.00	0.63	0.85	1.72	5.49	-0.03	-0.03	0.02	0.02
Congruency × Cue	0.00	0.01	0.50	0.81	1.01	4.33	-0.03	-0.02	0.04	0.04
LISAS										
Congruency	-46.2*	-38.2*	>0.99	1.00	3599	Inf	-67.7	-59.7	-17.0	-17.0
Cue	23.8*	33.9*	0.99	>0.99	69.0	1384	2.79	12.1	55.2	55.2
Congruency × Cue	-6.43	7.37	0.51	0.69	1.02	2.18	-36.5	-22.3	37.0	37.0
BIS										
Congruency	0.21*	0.26*	>0.99	1.00	1285	Inf	0.08	0.14	0.39	0.39
Cue	-0.17*	-0.12*	0.97	>0.99	28.6	304	-0.3.0	-0.25	0.01	0.01
Congruency × Cue	-0.03	0.04	0.50	0.66	1.01	1.90	-0.21	-0.15	0.22	0.22

Note. p_b = probability that the effect is in the noted (*b*) direction; ER_b = evidence ratio for the noted (*b*) direction; LISAS = Linear integrated speed-accuracy score (Vandierendonck, 2017); BIS = Balanced integration score (Liesefeld & Janczyk, 2019).

* $p_b > .95$

Figure F1

Residual Distribution of the Linear Mixed Model for Each of the Dependent Variable

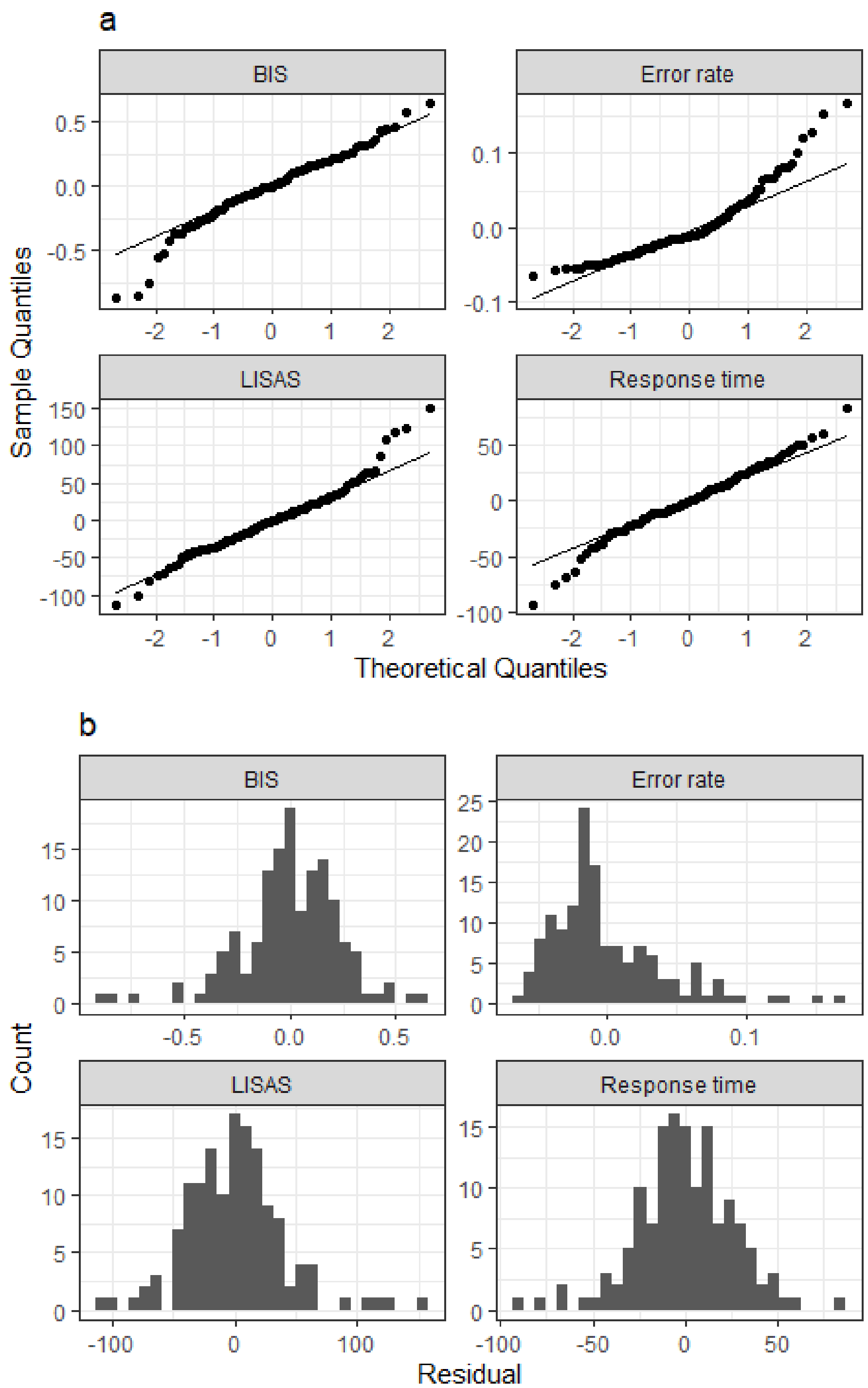


Table F7

Analysis of Variance With Outliers

Predictors	$F(1,25)$	p	η_p^2
Response time			
Congruency	30.0	< .001	0.546
Cue	38.8	< .001	0.608
Congruency × Cue	0.07	.797	0.003
Error rate			
Congruency	12.3	.002	0.329
Cue	0.43	.518	0.017
Congruency × Cue	0.02	.898	0.001
LISAS			
Congruency	37.8	< .001	0.602
Cue	24.5	< .001	0.495
Congruency × Cue	< 0.01	.985	0.000
BIS			
Congruency	35.1	< .001	0.584
Cue	15.7	< .001	0.385
Congruency × Cue	0.01	.924	0.000

Note. LISAS = Linear integrated speed-accuracy score (Vandierendonck, 2017); BIS = Balanced integration score (Liesefeld & Janczyk, 2019).

Table F8

Bayesian Mixed Model With Outliers Removed

				HDI	
Predictors	*b*	p_b	ER_b	Low	High
Response time					
Congruency	-20.8*	>0.99	449	-34.7	-6.95
Cue	37.5*	1.00	Inf	23.5	51.6
Congruency × Cue	-2.63	0.60	1.50	-22.6	17.4
Error rate					
Congruency	-0.02*	0.99	83.1	-0.04	0.00
Cue	0.00	0.72	2.62	-0.02	0.01
Congruency × Cue	0.00	0.55	1.23	-0.02	0.03
LISAS					
Congruency	-38.4*	>0.99	8999	-57.3	-19.4
Cue	35.6*	>0.99	17999	16.6	54.2
Congruency × Cue	-0.20	0.51	1.03	-26.9	26.5
BIS					
Congruency	0.20*	>0.99	5999	0.10	0.32
Cue	-0.18*	>0.99	946	-0.28	-0.07
Congruency × Cue	0.00	0.530	1.13	-0.15	0.16

Note. p_b = probability that the effect is in the noted (*b*) direction; ER_b = evidence ratio for the noted (*b*) direction; LISAS = Linear integrated speed-accuracy score (Vandierendonck, 2017); BIS = Balanced integration score (Liesefeld & Janczyk, 2019).

* $p_b > .95$

Appendix G

Experiment 2 Exploratory Analyses Tables and Figures

Table G1

Coefficients for the Three-way Interaction Model for Response Time

	Frequentist				Bayesian				
								HDI	
Predictors	β	*df*	*t*	*p*	*b*	p_b	ER_b	Low	High
Congruency	-48.1	2292	-3.77	< .001	-50.5*	1.00	Inf	-77.2	-23.2
Cue	33.0	2292	2.40	.016	32.7*	.987	74.9	3.82	61.5
Time	-55.8	2293	-3.59	< .001	-2.55*	>.999	8999	-3.95	-1.17
Congruency × Cue	17.7	2292	0.92	.359	20.0	.835	5.06	-20.1	60.6
Congruency × Time	48.6	2292	2.20	.028	2.23*	.987	75.6	0.25	4.18
Cue × Time	3.26	2293	0.14	.890	0.16	.563	1.29	-1.95	2.3
Congruency × Cue × Time	-45.0	2293	-1.35	.176	-2.07	.913	10.5	-5.06	0.89
Model fit									
Sigma (subject)					118*	1.00	Inf	115	122
R^2					.378			.353	.402
LOOIC					28935			SE = 101	

Note. The Time (block) variable is scaled to range from 0 to 1. Thus, the three-way interaction coefficient indicates the adjustment of the Congruency × Cue at the end of the experiment in contrast to the start of the experiment. p_b = probability that the effect is in the noted (*b*) direction; ER_b = evidence ratio for the noted (*b*) direction.

* p_b > .95

Table G2

Coefficients for the Three-way Interaction Model for Error Rate

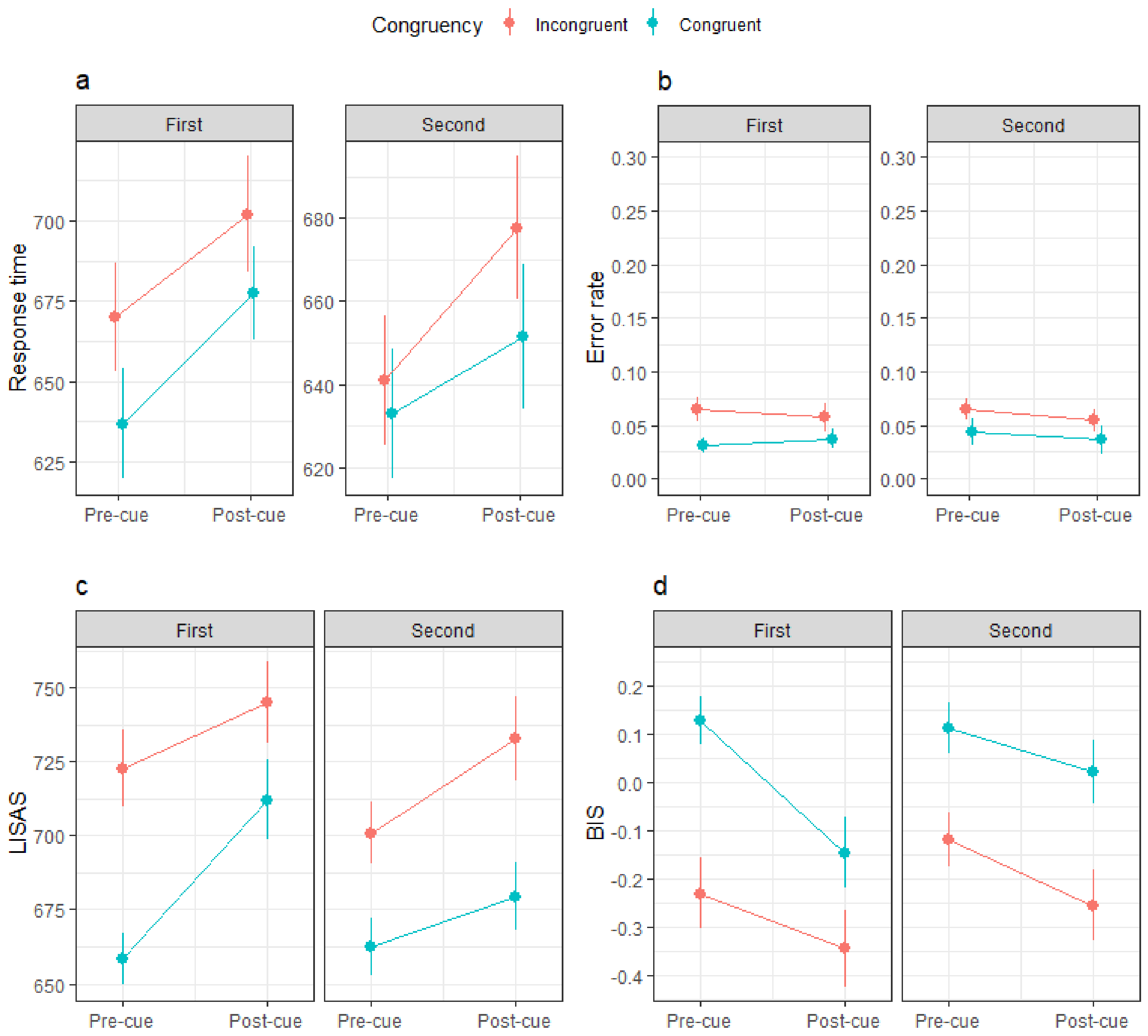
	Frequentist				Bayesian				
								HDI	
Predictors	β	*df*	*t*	*p*	*b*	p_b	ER_b	Low	High
Congruency	-0.05	2293	-3.00	.003	-0.06*	.998	513	-0.09	-0.02
Cue	-0.03	2296	-1.55	.121	-0.03	.937	14.8	-0.07	0.01
Time	-0.02	2300	-0.94	.346	0.00	.823	4.66	0.00	0.00
Congruency × Cue	0.06	2296	2.07	.038	0.06*	.979	46.0	0.00	0.12
Congruency × Time	0.04	2294	1.25	.211	0.00	.891	8.20	0.00	0.00
Cue × Time	0.04	2297	1.15	.251	0.00	.875	6.98	0.00	0.00
Congruency × Cue × Time	-0.08	2298	-1.70	.089	0.00*	.952	19.9	-0.01	0.00
Model fit									
Sigma (subject)					0.17*	1.00	Inf	0.16	0.17
R^2					.041			.025	.058
LOOIC					-1689			SE = 209	

Note. The Time (block) variable is scaled to range from 0 to 1. Thus, the three-way interaction coefficient indicates the adjustment of the Congruency × Cue at the end of the experiment in contrast to the start of the experiment. p_b = probability that the effect is in the noted (*b*) direction; ER_b = evidence ratio for the noted (*b*) direction.

* $p_b > .95$

Figure G1

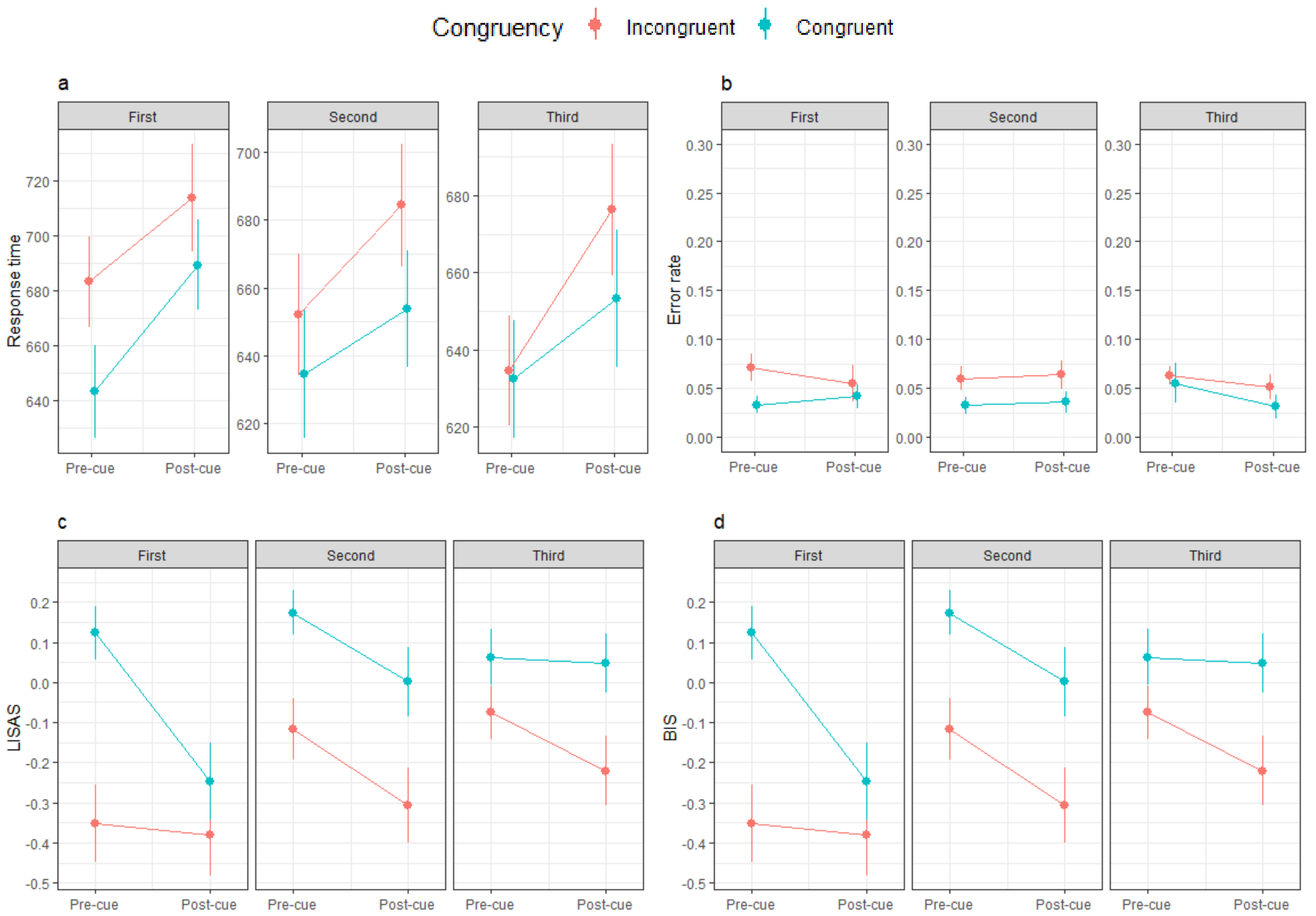
Experiment 2 Split in two Parts for (a) Response Time, (b) Error Rate, and the (c) LISAS and the (d) BIS



Note. The plots are split by two experiment lengths. The First correspond to the first half of the experiment run (i.e., 12 blocks), and the Second correspond to the last half of the experiment. Only the (a) response time appear to indicate the hypothesized two-way interaction in the Second subplot. That is, a reduced congruency effect during the pre-cue but a presented congruency effect during the post-cue. LISAS = Linear integrated speed-accuracy score (Vandierendonck, 2017). BIS = Balanced integration score (Liesefeld & Janczyk, 2019).

Figure G2

The Two-way Interaction Between (a) Response Time, (b) Error Rate, (c) LISAS and the (d) BIS, Split by Three Experiment Parts



Note. The subplots are split by three experiment lengths. The First correspond to the first 1/3 of the experiment run (i.e., 8 blocks), the Second to the next 1/3 of the experiment and the Third to the last 1/3 of the experiment. On the Third subplot, the hypothesized two-way interaction appears to be present – most evident for (a) response time. That is, a reduced congruency effect during the pre-cue but a presented congruency effect during the post-cue. LISAS = Linear integrated speed-accuracy score (Vandierendonck, 2017). BIS = Balanced integration score (Liesefeld & Janczyk, 2019).

Appendix H

Investigating the Effect of Time in Abrahamse et al. (2022) Experiment 3

In Abrahamse et al.'s (2022) third experiment, they investigated whether cancelling the inducer instructions using a cue could remove or attenuate the congruency effect on the diagnostic trials after the cue. Their results revealed that the congruency effect was presented even after the cue. Despite the results, the authors did not investigate whether the attenuation of the congruency effect after the cancellation cue interacted with time. Considering this, I attempted to investigate whether the congruency effect would interact with time in the cancellation condition run type. My analysis did not reveal a significant interaction with time (see Table H1). However, there was lacking information regarding the coding of the "run_type" variable. Additionally, I did not have an opportunity to contact the author regarding this variable. Nevertheless, none of the run types indicated a significant interaction with Time, suggesting that the congruency effect in Abrahamse et al.'s experiment 3 did not appear to change with time.

Table H1

Regression Analysis for the Congruency x Time Interaction

	Frequentist			Bayesian				
							HDI	
Predictor	β	*SE*	*t*	*b*	p_b	ER_b	Low	High
Congruency	-7.50	6.49	-1.16	-7.57	.881	7.39	-20.2	5.05
Time	-39.3	7.64	-5.15	-39.4*	1.00	Inf	-54.2	-24.6
Congruency × Time	-5.22	10.7	-0.49	-5.11	.686	2.19	-25.7	15.5

Note. p_b = probability that the effect is in the noted (*b*) direction; ER_b = evidence ratio for the noted (*b*) direction.

* $p_b > .95$