Cognitive Load Impairs Professional Scepticism in Decision-Making: The Mitigating Role of Default Nudges

Mercede Erfanian [1*], Luc Meunier [2], Jean-Francois Gajewski [3]

[1] ESSCA School of Management, Lyon, France
[2] ESSCA School of Management, Aix-En-Provence, France
[3] University of Lyon, Jean Moulin, iaelyon, Magellan, Lyon, France

Abbreviated title: Cognitive load and default nudges role in scepticism

***Corresponding author:** Mercede Erfanian
mercedeh.ERFANIANGHASAB@essca.fr

Number of Figures: 3
Number of Tables: 1
Number of words for Abstract: 243
Number of words for Introduction: 1671
Number of words for Discussion: 2205

ORCID IDs:

Mercede Erfanian https://orcid.org/0000-0001-9253-4162

Luc Meunier https://orcid.org/0000-0002-9139-8530

Jean-Francois Gajewski https://orcid.org/0000-0003-2294-5761

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

The REMOTAUDIT project has received funding from the French Research Agency (Agence Nationale de la Recherche (ANR)), grant ANR-21-CE26-0012-02. We are thankful to the ANR team members in particular Marco Heimann for his useful feedback on the experimental design. We also thank Ehsan Eqlimi for helping with analysis codes and feedback on the draft.

Data Availability Statement

All raw and processed data are available at (https://osf.io/s4tc8/). The MATLAB code for statistical analyses, and visualizations will be made available alongside the experimental data upon publication at (https://github.com). An example video of the experimental procedure is available at the following link (https://vimeo.com).

Author Contributions

ME conceptualized the study and developed the methodology with input from LM and JFG. ME conducted data collection, performed formal analysis, and curated the data. ME drafted the original manuscript, with editing conducted collaboratively and technical feedback by LM and JFG. LM and JFG supervised the project. Additionally, LM and JFG oversaw project administration and supported funding acquisition efforts.

Abstract

Cognitive overload may compromise critical competencies essential for effective auditing, including fraud detection and risk assessment. Auditors, tasked with processing extensive information under time constraints, may experience difficulties in maintaining professional scepticism, a fundamental mechanism that supports these core evaluative functions. Default nudges or pre-selected options, subtly shape choices with the premise of easing cognitive load and ultimately enhancing audit performance. This study examined how cognitive load and default nudges shape decision-making-based professional scepticism. In three online experiments, we investigated how default nudges influence professional scepticism under varying cognitive loads among auditors. Experiment 1 validated a cognitive load manipulation (minimal, low, medium, high) using reaction time and accuracy, selecting low and high load conditions for further study. Experiment 2 confirmed load-dependent reaction time increases, with audit task accuracy remaining stable. Experiment 3a tested the effect of default nudge on professional scepticism under varying levels of cognitive load, revealing that nudging accelerated responses and improved audit accuracy, but only under cognitive load. Experiment 3b explored default-setting-response alignment, showing that nudging toward considering an accounting item as "non-aggressive" (e.g., not indicative of fraud), the less sceptical and most common case, enhanced accuracy, whereas nudging toward considering the item "aggressive", reduced it. These findings highlight how cognitive load disrupts scepticism and suggest that strategically placed nudges can enhance audit decision-making based professional scepticism under pressure. Insights from this study have direct implications for high-stakes professions, where optimized decision environments could support critical judgment under cognitive strain.

Keywords: Decision-making, professional scepticism, working memory, cognitive load, default nudges, audit

Cognitive Load Impairs Professional Scepticism in Decision-Making: The Mitigating Role of Default Nudges

Introduction

The human brain, accounting for only approximately 2% of total body weight, is a metabolically demanding organ, consuming around 20% of the body energy at rest (Raichle & Gusnard, 2002). Despite its high energy requirements, it is remarkably efficient, employing sophisticated mechanisms to minimize metabolic costs while maintaining optimal functionality (Li & van Rossum, 2020; Padamsey & Rochefort, 2023; Raichle & Gusnard, 2002; Sterling & Laughlin, 2015). This metabolic efficiency has significant implications for the brain executive functions (EF) (Baumeister, 2001; Gailliot, 2008) particularly decision-making (DM) (Wiehler et al., 2022). DM represents a complex cognitive ability that integrates and relies upon other EFs such as working memory (WM), cognitive control, and cognitive flexibility (Botvinick et al., 2004; Mansouri et al., 2017).

Scepticism, defined as the disposition to question assumptions and critically evaluate evidence, is a pivotal element of superior DM performance. It is a cognitively demanding process that necessitates considerable mental effort and WM resources (Evans & Stanovich, 2013; Normand, 2008). It engages neural circuits with heightened activity observed in but not limited to the anterior cingulate cortex (ACC) (Kashay et al., 2022; Pennycook et al., 2019). This region supports error detection, conflict monitoring, and the adjustment of cognitive strategies, all of which are energetically costly but crucial for adaptive scepticism (Botvinick et al., 1999; Botvinick, 2007; Botvinick et al., 2001; Botvinick et al., 2004; Carter et al., 1998; Kerns et al., 2004; Mansouri et al., 2017). Prolonged engagement in sceptical reasoning often leads to cognitive fatigue, characterized by diminished attentional control, slower information processing, and reduced cognitive flexibility (Hagger et al., 2010). Despite these challenges, maintaining a high level of scepticism is essential for effective DM and performance in certain high-stakes tasks and professions such as accounting, finance, and audit (Hurtt, 2010; Hurtt et al., 2013).

Auditing is a structured and methodical process designed to gather sufficient and appropriate evidence to determine whether organizations comply with legal requirements (compliance audit) and whether their financial reporting adheres to the applicable accounting principles and standards (financial audit). Additionally, auditing includes an independent, objective, and systematic evaluation of an organization operations, programs, activities, or public functions, intending to assess the extent to which available resources are utilized efficiently, effectively, and economically (performance audit) (González-Díaz & García-Fernández, 2018).

In the audit, scepticism or *professional scepticism* serves as a cornerstone for fraud detection, risk assessment, and the enhancement of audit quality (Gajewski, Heimann, Léger, et al., 2024; PCAOB, 2022, 2023). Professional scepticism is best defined as *'an attitude that includes a questioning mind, being alert to conditions which may indicate possible misstatement due to error or fraud, and a critical assessment of evidence'* (Gajewski, Heimann, Léger, et al., 2024). Auditors and financial analysts are required to critically evaluate financial statements and identify anomalies within a limited timeframe that may indicate misconduct or systemic risk (Asare et al., 2000; Board, 2012; Johari et al., 2019; Lambert et al., 2017; McDaniel, 1990; Otley & Pierce, 1996; Svanström, 2016). Low professional scepticism in this context can have detrimental consequences, such as financial losses or large-scale economic crises (Hurtt, 2010; Nelson, 2009). Given the cognitive demands associated with maintaining professional scepticism, it is crucial to identify factors that reduce professional scepticism and negatively impact DM, as well as strategies to mitigate fatigue-inducing effects in cognitively demanding such a high-stakes profession.

The modernized nature of contemporary life has led to an unprecedented surge in the volume of information individuals are exposed to within increasingly condensed periods of time (Rutkowski & Saunders, 2018), reaching approximately 74 GB per day (Bohn & Short, 2012). This constant barrage of information often surpasses the processing capacity of the human brain, which operates with finite cognitive resources. As competing information increases, attentional demands exceed available capacity, thereby attenuating EF, a principle central to the Time-Based Resource-Sharing (TBRS) model (Barrouillet et al., 2004; Barrouillet et al., 2007). This model assumes that both storage limitations and the processing (e.g., DM) demands compete for attentional resources (Barrouillet & Camos, 2012). As a result, the simultaneous demands on cognitive resources, in particular WM, can lead to cognitive overload, a state in which the cognitive demands exceed an individual's capacity (Sweller, 1988). This state is associated with increased neural activity and elevated metabolic demands (neurovascular coupling, NVC), contributing to high brain energy consumption (Backs & Seljos, 1994; Csipo et al., 2021).

There is evidence to suggest that cognitive overload alters executive functioning including inhibitory control (Barzykowski & Niedzwienska, 2018; Mazzoni, 2019; Vannucci et al., 2015), problem-solving (Ninomiya et al., 2024; Sweller, 1988), and DM (Ordali et al., 2024) by reducing the ability to process information efficiently and critically (Sweller, 1988). This impact is particularly pronounced in complex tasks requiring vigilance, as professional scepticism in the context of DM does. Additionally, neural fatigue as a result of highly cognitively demanding tasks depletes prefrontal cortical resources, which are critical for DM (Sridharan et al., 2008). Consequently, individuals experiencing cognitive overload may rely more on heuristic-based or automatic processing, which undermines sceptical inquiry and increases susceptibility to biases (Evans & Curtis-Holmes, 2005; Gigerenzer, 2008; Van Gestel et al., 2021). In the audit, this decline in professional scepticism due to cognitive overload can result in suboptimal evaluations of risk, failure to detect fraud, and overall compromised audit quality (Austin, 2023; Bedeir, 2024; Nelson, 2009; Rose, 2007).

Default options are pre-set courses of action that take effect if nothing is specified by the decision maker (Thaler & Sunstein, 2008). They involve presenting a pre-selected option while allowing individuals the opportunity to opt out (Gajewski, Heimann, Meunier, et al., 2024). Default nudges are widely recognized as highly potent instruments for influencing behaviour across different disciplines, particularly in behavioural economics, by subtly modifying the choice architecture (Mertens et al., 2022; Münscher et al., 2016). Their mechanism of action is generally thought to originate in the status quo bias, the tendency of people to stick with the current state of affairs, which is often the default option (Gajewski et al., 2022; Meunier & Richit, 2024). They have been shown to influence a range of behaviours, including risk-taking (Meunier et al., 2024), socially responsible investment (Gajewski et al., 2022; Meunier & Richit, 2024), and financial DM (Gajewski, Heimann, Meunier, et al., 2024), among others. By simplifying the DM process and reducing the cognitive effort required to evaluate multiple alternatives, default nudges effectively mitigate the adverse effects of cognitive overload on decision quality (Kool & Botvinick, 2018; Meunier et al., 2024; Thaler & Sunstein, 2008). For instance, Gigerenzer and Gaissmaier (Gigerenzer & Gaissmaier, 2011) highlight that simple heuristics, like defaults, align with the brain tendency to conserve cognitive energy while enhancing rational outcomes.

Auditors often face high-pressure situations with extensive information and tight deadlines, which can conceivably impair their ability to detect fraud or evaluate risk effectively (Lopez & Greenwald, 2022; Nelson, 2009; Persellin et al., 2019). Kunda and Oleson (Kunda & Oleson, 1995), as an example, emphasize that structured decision environments, including defaults, promote analytical thinking and reduce biases, particularly in high-stakes scenarios. Similarly, Thaler and Sunstein (Thaler & Sunstein, 2008) argue that nudges are particularly effective in professional settings where individuals are prone to decision fatigue.

Hence, in an online study, we examine how variations in cognitive load affect professional scepticism in auditors. Additionally, we investigate the impact of default nudges as mechanisms for influencing professional scepticism. During the delay period of a short-term memory task, participants engaged in an audit task specifically designed to probe their professional scepticism as a proxy of optimal DM. Subsequently, we assessed their objective scepticism. This approach allowed us to characterize the interaction between default nudges, task-induced cognitive demands, and the manifestation of sceptical tendencies.

In experiment 1, we examine the effects of task manipulation on cognitive load (‘minimal load’, ‘low load’, ‘medium load’, and ‘high load’), imposed through the Dot Memory Task (DMT) and indexed by accuracy (%) and reaction time (RT, milliseconds), to verify that these conditions are behaviourally dissociable. This step is essential to confirm that cognitive load is being effectively and systematically manipulated. During the retention interval of a short-term DMT, participants engage in four basic mathematical operations. Cognitive load is modulated by varying the number of dots and the size of the grid (see Methods). We hypothesize that the high load condition will significantly modulate RTs and accuracy relative to the other conditions (H1). As hypothesis, RTs were significantly longer in proportion to the increase in load, whereas accuracy reached its lowest level under the high-load condition.

In experiment 2, we explore the impact of cognitive load (‘No Load’ (NL), ‘Low Load’ (LL), and ‘High Load’ (HL)), induced by taxing WM, on performance in an audit task designed to assess professional scepticism. Performance was measured through RT and accuracy. We hypothesize that increasing cognitive load will lead to longer RTs in the audit task, accompanied by reduced accuracy. In other words, under high cognitive load, participants are expected to perform worse compared to conditions with NL or LL, and worse under LL compared to NL (H2). While audit task accuracy remained stable despite increasing cognitive load, RTs significantly increased.

In experiment 3a, we examine the influence of default nudges on performance in the audit task across three levels of cognitive load, comparing a group exposed to the nudges with a group that is not. We predict that participants exposed to default nudges will demonstrate faster RTs and greater accuracy in the audit task relative to those who are not exposed (H3). In experiment 3b, we extend this investigation to determine the optimal placement of the default nudges, testing whether they should be aligned with the option associated with a higher or lower probability of correct responses. We hypothesize that default nudges will decrease RTs and enhance accuracy when placed on the option with a higher probability of correctness (H4). The findings confirmed our hypothesis: default nudges, when applied to the option with a higher probability of a correct response, enhanced both RTs (as reflected in faster responses) and accuracy (as reflected in higher correctness) under load conditions.

Methods

Participants were recruited from Prolific (www.prolific.com), an online study pre-registered subject pool. This platform guarantees the recruitment of a homogenous sample that fits well with the specified inclusion and exclusion criteria. Once the inclusion criteria were established, the study advertisements were exclusively visible to participants meeting these eligibility requirements. Within the advertisement on Prolific, participants were provided with a brief overview of the study and the link to the online experiment. The Gorilla Experiment Builder platform (www.gorilla.sc) (Anwyl-Irvine et al., 2020) was used to build and host the online experiments. All procedures were approved by the local ethics committee of the School of Management ESSCA (Reference No. Avis 7/2024, dated 19 August 2024). All participants provided electronic informed written consent prior to the experiment. The study was registered on OSF (https://doi.org/10.17605/OSF.IO/AGMWK).

Prior to the main experiment hosted on Gorilla, registered participants were informed about the study requirements, including being in a quiet and distraction-free environment. A brief description of the study objectives was provided, along with details on participants' roles during the experiment, the estimated time required to complete the tasks, the compensation they would receive upon successful completion, and the system requirements (e.g., using Google Chrome on a desktop or laptop computer). It was explicitly stated that participants would not be compensated if they failed to meet the system requirements, did not complete the experiment successfully, or achieved a performance level below 67%. Participants then were directed to the consent form before reporting their audit experience, which ranged from 'less than 1 year' to 'more than 10 years' (only presented to participants of experiment 2 onwards). Following this, they advanced to the main experiment.

After completing the main experiment, participants who took part in experiment 2 and beyond were asked to rate the extent to which they perceived the "pre-selected" options or default nudges in the test as influencing their decisions regarding each piece of evidence. Responses were recorded on a 6-point Likert scale ranging from "not at all (1)" to "extremely (6) **(Table 1)**. The experiment ended with a debriefing.

Inclusion criteria

Inclusion criteria for all experiments were a) being 18 years old or more and under 60 years old, b), having no dyslexia or dyscalculia, c) having normal vision with no colour-blindedness, and for experiment 2 onwards d) being either a professional auditor or student of finance or business with knowledge of auditing. The age range was selected to preserve consistency in cognitive functioning (Salthouse, 2009; Salthouse et al., 2008).

Participants

A total of 299 participants initially took part in the study (38 in experiment 1, 137 in experiment 2, 51 in experiment 3a, and 73 in experiment 3b). Across all experiments, data from 239 out of 299 participants (experiments 1 to 3b) were included in the analyses, including *only* those who passed the attention check (See Experimental measures). It is important to note that the number of participants varied across different phases of the study, and data from participants who did not complete all study phases were retained in the analysis (See Data analysis) **(Table 1).**

Table 1. *The table below presents the number (N) and percentage (%) of professional auditors based on their years of experience in auditing, as well as the extent to which the pre-selected option influenced their DM-based professional scepticism.*

Category	*N*	Percentage (%)
Years of Experience		
Less than 1 year	83	31.3
1–5 years	136	51.3
5–10 years	28	10.6

More than 10 years	8	3.0
Other	9	3.4
Influence of Pre-Selected Options		
Not at all	17	14.9
A little	22	19.3
Somewhat	29	25.4
Quite a bit	22	19.3
A lot	14	12.3
Extremely	10	8.8

Experiment 1. 35 participants (out of 38) completed the online experiment (age range: 18–60 years, M = 38.4 ± 11.87; 48.57% female). The study subjects received a fair-based reimbursement for the time they spent on the online experiment (recommended by Prolific £9 (€10.8) per hour – received €0.75 for ~ 5 min). A priori power analysis was conducted to determine the appropriate sample size and establish adequate statistical power for detecting effects. The analysis revealed that 34 subjects would provide 81% power to detect a medium effect size (*d* = 0.5) across the experimental conditions with an alpha of 0.05 (Cohen, 1988).

Experiment 2. The final analysis included data from 108 participants out of 137 (aged 18 to 58 years, mean age = 35.2 ± 9.81 years, 46.59 % female) as not all met the attention check criteria. A priori power analysis showed that 108 subjects would provide 99% power to detect a medium effect size (*d* = 0.5) across the experimental conditions with an alpha of 0.05.

Experiment 3a, 3b. Two separate groups of participants were recruited (N = 38 out of 51 for nudge – Disagree (nudge (D)), and (N = 58 out of 73 for nudge – Agree (nudge (A)), female: 50 %, mean age: 40.46 ± 7.58 years, power = 93%, with a large effect size of *d* = 0.8, alpha = 0.05. For experiments 2,3a and 3b, participants received €2.15 provided for ~15 min of participation. Additionally, to enhance performance on the online tasks (Bianco et al., 2021), participants were offered a performance-based monetary bonus of €0.05 for each correct response on both the DMTs and audit task.

Stimuli and procedure

Experiment 1. Figure 1 shows an overview of the experimental design. The Dot Memory Task (DMT) (Bethell-Fox & Shepard, 1988) was used to modulating cognitive load by manipulating the complexity of visual memory retention across four conditions (‘minimal load’, ‘low load’, ‘medium load’, and ‘high load’) (Miyake et al., 2001). In this task, participants were presented with a cue trial in which the dot pattern appeared in the grid for a pre-determined duration (1s). Following the presentation of the grid, it disappeared, marking the onset of a brief retention interval intended to minimize subvocal rehearsal of the dot positions and encourage reliance on WM. During this interval, participants engaged in a secondary task involving one of the four basic mathematical operations (e.g., subtraction), with responses constrained to values less than or equal to 10. They were required to complete this task within a 5s time window, indicated by a countdown timer displayed in the top-right corner of the screen. Immediate visual feedback was provided after each trial: a red cross for incorrect and a green checkmark for correct responses. This task was intentionally designed to be readily manageable for participants while simultaneously imposing cognitive demands to tax WM resources. The following participants started with a 3-trial practice block. Participants were required to achieve an accuracy of 67% before proceeding to the main task; otherwise, the task was repeated until they met this criterion. Participants then completed four experimental blocks, each consisting of 12 randomly presented trials, resulting in a total of 48 trials across the experiment. The four blocks corresponded to varying levels of cognitive load that were defined the number pf dots and size of the grid: minimum (2x2 grid with 2 dots), low (3x3 grid with 2 dots), medium (3x3 grid with 3 dots), and high (4x4 grid with 4 dots). These conditions were counterbalanced across participants to minimize order effects **(Figure 1).**

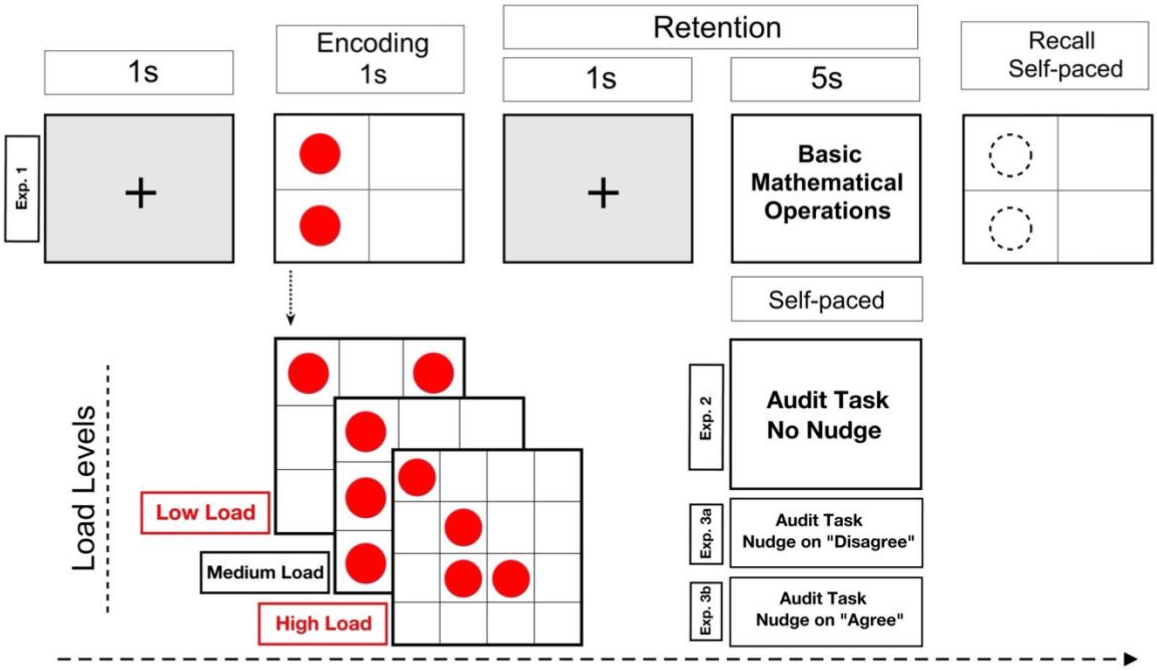


Figure 1. *The sequence and timing of DMT (e.g., medium load). In experiment 1, the task began with the presentation of a fixation cross for 1s, followed by a 1s encoding period during which participants were shown a dot pattern. This was followed by a 1s retention interval and a 5s interval featuring one of the four basic mathematical operations (e.g., 8 - 4=?), designed to prevent rehearsal. In the recall phase, participants reconstructed the dot pattern within a self-paced response window, testing their WM under varying cognitive load conditions. For subsequent experiments, cognitive load was manipulated using low and high load conditions, marked in red. In experiment 2, the sequence began the same as experiment 1, followed by a Phillips task piece. Participants decided whether the stimulus indicated 'aggressive auditing' by clicking "Yes" or "No" in a self-paced manner with no time limit. In experiment 3a, participants were presented with a pre-selected response to determining whether the stimulus indicated 'aggressive auditing'. If participants disagreed with the pre-selected response, they were instructed to simply click "Continue." Conversely, if they agreed, participants were required to drag a checkmark icon initially placed on the "disagree" option to "agree" before clicking "Continue." In experiment 3b, all the steps were identical except the pre-selected response was placed on "agree". Participants were allowed to complete this phase at their own pace. During the recall phase, participants reconstructed the dot pattern within a self-paced response window.*

Experiment 2. The online experiment replicated the structure of the prior experimental design with a key modification to integrate audit-related judgments **(Figure 1)**. Additionally, in place of the original four difficulty levels, we employed NL (No Load) condition, alongside the LL and HL conditions from experiment 1. Prior to the task, participants were presented with a detailed audit scenario, which served as the context for evaluating subsequent audit pieces of evidence. The task began with a fixation cross, followed by the encoding of a randomly generated pattern of dots, and was then followed by a retention interval. During designated intervals within the experiment, participants were presented with individual audit pieces of evidence sequentially. For each piece of evidence, they were required to make a judgment based on the initial audit scenario, answering the question: "Is this piece indicative of aggressive reporting?" Participants responded at their own pace by selecting either "Yes" or "No." In the NL condition, participants viewed the evidence and responded immediately after the first fixation cross, with no DMT included. The remainder of the experiment, including the core task structure and timing parameters, remained identical to the previous experimental protocol, ensuring comparability between conditions.

Experiment 3a and 3b. The setup of the DMT and audit tasks were designed to replicate experiment 2. However, this experiment incorporated a default nudge, where instead of actively selecting “Yes” or “No”, a response was pre-selected for the participants. Specifically, two options were provided: “Agree” and “Disagree”, with the pre-selected response set to the “Disagree” option in experiment 3a. The rationale for placing the default nudge on the “Disagree” option is based on its higher likelihood of being the appropriate response. Specifically, approximately 58.33% of the audit-related items (out of 12 in each block, including two attention check items and 10 audit pieces) were classified as non-aggressive, necessitating “Disagree” as the correct response. While the general structure of the experiment remained consistent with the previous iteration, participants were instructed to carefully read the audit pieces. If they disagreed that the piece was indicative of aggressive auditing (or agreed with the pre-selected response or default nudges), they simply clicked "Continue." However, if they agreed that the statement, they were required to drag the checkmark initially placed on the "Disagree" option to the "Agree" option before proceeding by clicking “Continue”. Experiment 3b was identical to experiment 3a, except that the pre-selected option was initially placed on “Agree”; all other aspects remained the same.

Experimental measures

Phillips audit task. The Phillips audit task is a widely used tool in the field of auditing (Phillips, 1999). In this task, participants are presented with background information describing audit scenarios involving clients who manufacture and market water, electricity, and natural gas meters. The information notes that the metering industry has experienced recent growth due to technological innovations and that the client company is a growing entity with increasing revenue, income, and assets. After reading this background information, participants proceed to a self-paced review of audit evidence consisting of 45 statements, each related to one of 14 financial statement accounts. In this phase, participants assess each statement to determine whether it indicates ‘Aggressive Reporting’ by selecting either a ‘Yes’ or ‘No’ response. In our study, we included 30 statements, with 21 classified as non-aggressive and 9 as aggressive, distributed equally across three separate blocks, making 7 non-aggressive and 3 aggressive statements per block.

Attention check. To confirm that participants’ responses were genuine and that they maintained attentiveness throughout the experiment, we incorporated attention check questions into the task starting from experiment 2 onward. These attention checks were framed within the context of audit-related pieces but required participants to perform straightforward arithmetic calculations. For example, one attention check presented the following statement: “Meter-Tek purchased 300 units in January and sold 150 units in February, leaving a remaining inventory of 200 units.” Participants were then asked, “Is this correct?”. The attention check trials were designed to be simple and unambiguous. These checks were incorporated pseudo-randomly into the task, ensuring that their placement was balanced across participants. In total, six attention check questions were included in the study, with two checks placed randomly within each block.

Objective scepticism. Consistent with (Gajewski, Heimann, Léger, et al., 2024), we tallied objective scepticism. The frequency or proportion of detected aggressive items serves as an index of objective scepticism, with higher values indicating greater levels of sceptical tendencies.

Data analysis

In all experiments, RTs (in milliseconds (ms)) and accuracy (in percentage (%)) were computed as dependent measures indexing cognitive load in the DMT and DM-based professional scepticism in the audit task. RTs during the recall phase were operationalized as the total time elapsed from the presentation of the empty recall matrix to the participant's final dot placement. This measure reflects the full temporal span of the recall process, including memory retrieval, spatial DM, and motor execution (Sliwinski et al., 2018). Participants were not permitted to modify their responses once the final dot was placed. RTs were averaged across all trials per participant and then per condition (e.g., LL).

Accuracy was scored dichotomously: a score of 1 was assigned when all target dots were placed in their correct grid locations, and 0 otherwise, yielding a stringent measure of recall fidelity. Accuracy was calculated as the proportion of correctly completed trials per participant, then averaged across participants and multiplied by 100 to obtain percentage accuracy.

We adopted a lenient approach to outliers. Since the goal was to examine the effect of cognitive load, as a function of processing, on pattern retention in WM without introducing time constraints (Hu et al., 2024), the timing was not manipulated in the task. Thus, participants reconstructed the pattern at their own pace. Given that WM was already taxed by the concurrent audit task, they were expected to reconstruct the pattern as quickly as possible to minimize memory lapse. This approach aimed to isolate cognitive load effects while reducing time pressure.

Thus, both RTs and accuracy outliers were retained. RTs naturally vary due to individual differences in cognitive strategies, attention, and prior experience, all of which are integral to learning and memory. Removing long RTs could eliminate meaningful strategic differences in recall, as some participants may take longer to process and reconstruct patterns in ecologically valid ways (Miller, 1991). Similarly, accuracy variations arise from cognitive factors like momentary lapses in attention rather than random noise. Removing accuracy outliers could exclude participants whose performance, while seemingly suboptimal, remains informative for understanding cognitive processes (Miller, 2023). Extreme RTs and inaccuracies provide information about individual differences, making their retention essential for a comprehensive analysis of memory recall under no time pressure. Nonetheless, we also report all analyses of experiments 1, 2, 3a, and 3b with outliers (exceeding ±3 SD from a participant's mean) in RTs and accuracy removed (see Supplementary Materials). Outlier RTs were identified and removed per participant, while outlier accuracy rates were detected and excluded at the group level across participants, prior to statistical analyses. Subsequently, only participants with more than 50% missing trials (including DMT and audit tasks) were removed from the dataset. In line with Prolific guidelines, participants who failed to meet the attention check criteria threshold (67%, corresponding to 2 out of 6 attention checks) were excluded from the analysis.

Experiment 1: An increase in cognitive load associates with longer RTs and reduced accuracy

Results

Figure 2 shows the behavioural performance including RTs **(Panel 1A)** and adjusted accuracy **(Panel 1B)** of DMT in experiment 1. We first computed accuracy scores such that they were adjusted to account for trials in which participants provided only correct responses to both DMT and mathematical operations (adjusted accuracy) (incl. correct hit) **(Panel 1B)**. This adjustment was implemented to ensure participants were actively engaged in the task while their WM capacity was being cognitively loaded. Then we calculated accuracy taking into account correct and incorrect mathematical operations (accuracy + math) (incl. correct and incorrect responses) **(Panel 1C)**. Finally, we calculated accuracy based on responses to DMT regardless of participants' responses to the mathematical operations (overall accuracy) (incl. miss, correct and incorrect responses) **(Panel 1C)**. The observed results remained consistent between adjusted accuracy, accuracy + math and overall accuracy.

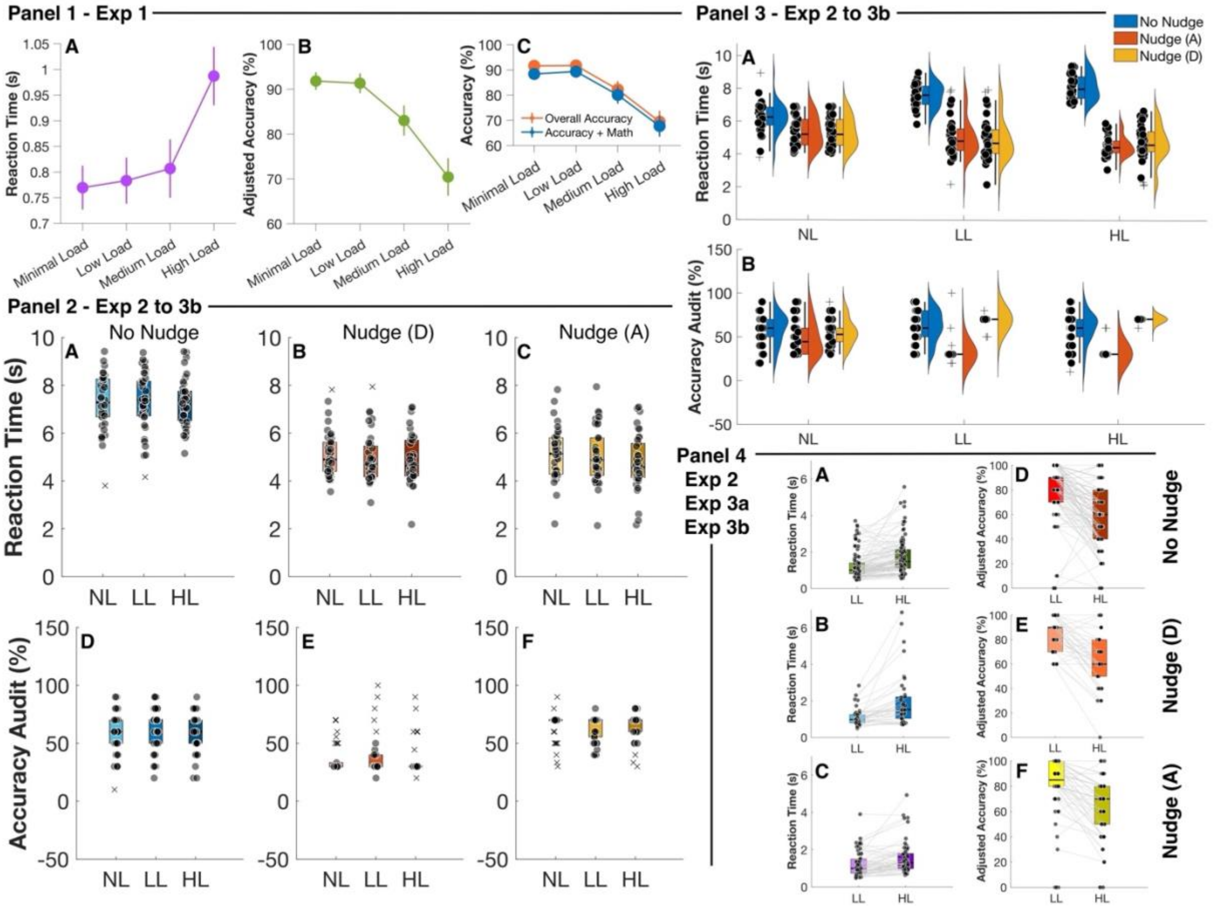


Figure 2. Panel ***1A*** illustrates DMT RTs, panel ***1B*** depicts adjusted accuracy. Panel ***1C*** shows overall accuracy and accuracy + math. The error bars represent ±1 SEM. Panel ***2A, B,*** and ***C*** depict the behavioural performance for the audit task, showing reaction time (s), and ***2D, E,*** and ***F*** show accuracy rate (%) in experiments 2, 3a, and 3b, respectively. Each whisker represents the interquartile range (IQR), the horizontal black line denotes the mean. Panel ***3A*** presents audit task reaction time and panel ***3B*** accuracy across load conditions between no nudge, nudge (D), and nudge (A) groups. The kernel density plots visualize the data distribution. Panels ***4A, B,*** and ***C*** illustrate the reaction time in the DMT per experimental group including no nudge, nudge (D) and nudge (A), and panels ***4D, E,*** and ***F*** show the accuracy of the audit task in no nudge, nudge (D) and nudge (A) groups. All circles represent individual data points. Abbreviations: NL (No Load), LL (Low Load), and HL (High Load) on x-axes.

A Friedman test yielded significant differences in RT across load conditions $\chi2(3) = 33$, $p < 0.001$ (**Panel 1A**). The Wilcoxon post-hoc analyses (with Bonferroni correction) indicated significant differences in rank RTs of high cognitive load with other groups: minimal load ($Z = -4.586$, $p < 0.001$), low load ($Z = -4.209$, $p < 0.001$), medium load ($Z = -4.143$, $p < 0.001$).

A similar trend was observed in accuracy (%) as a function of the task difficulty level (**Panel 1B**). A Friedman test confirmed differences across the load conditions $\chi2(3) = 25.58$, $p < 0.001$, with significant differences between high load and other groups through Wilcoxon post-hoc with Bonferroni correction: minimal load ($Z = 3.995$, $p < 0.001$), low load ($Z = 4.471$, $p < 0.001$), and medium load ($Z = 2.711$, $p < 0.001$) for adjusted accuracy **(Panel 1B).** For overall accuracy, the test also revealed differences between groups, $\chi2(3) = 28$, $p < 0.001$. Again, the high load condition appeared to be different from the minimal load (Z = 4.031, $p < 0.001$), low load (Z = 4.578, $p < 0.001$), and medium load (Z = 2.717, $p < 0.001$) **(Panel 1C).** Finally, for accuracy + math, differences between groups were observed, χ2(3) = 27.4, $p < 0.001$: with significant differences between high load and minimal load (Z = 3.88, $p < 0.001$), and low load (Z = 4.541, $p < 0.001$) **(Panel 1C).** These results suggest that the manipulation of cognitive load effectively created distinct levels of difficulty, as evidenced by significant differences in both RT and accuracy. Consequently, the low and high load conditions were selected to modulate cognitive load in the subsequent experiments.

Experiment 2: Cognitive load is attributed to prolonged RTs of scepticism task performance while does not affect the accuracy

Results

Rank RTs of DMT were computed for each participant as a function of load. A Wilcoxon signed-rank test with the within-subject factors of 'no nudge LL' and 'no nudge HL' on RT data showed Z = -4.533, $p < 0.001$ **(Panel 4A)**, and on accuracy rate Z = 5.089, $p < 0.001$ **(Panel 4D)**.

To test the effect of cognitive load on the audit task, we conducted the Friedman test with the within-subject factors ('NL', 'LL' and 'HL'), which yielded a significant main effect of load, $\chi^2(2) = 99.69$, $p < 0.001$. Post-hoc pairwise comparisons using the Wilcoxon signed-rank test revealed that RTs were significantly slower in the LL compared to the NL condition (Z = -7.14, $p < 0.001$), slower in the HL compared to the NL condition (Z = -8.406, $p < 0.001$), and slower in the HL compared to the LL condition (Z = -4.751, $p < 0.001$) **(Panel 2A).** Unlike RTs, the accuracy rate of the audit task did not vary as a function of cognitive load modulation across the levels, $\chi^2(2) = 1.53$, $p < 0.466$ **(Panel 2D) (Panel 3A and 3B).** No post hoc analysis was conducted. This finding indicates that although cognitive load manipulation reliably attenuates response speed in the audit task, it does not compromise response accuracy, indicating that performance precision is resilient to variations in cognitive demand.

Experiment 3a: Default nudges enhance response efficiency by decreasing RTs and improving accuracy under cognitive load

Results

Consistent with the experiment 1, the Wilcoxon signed-rank test with 'nudge (D) LL' and 'nudge (D) HL' as within-subject factors on RTs revealed Z = -4.793, $p < 0.001$ **(Panel 4B)**, and on accuracy rate Z = 4.32, $p < 0.001$ **(Panel 4E)**, confirming the cognitive load was also successfully manipulated in the nudge (D) group.

A two-way mixed ANOVA with Greenhouse-Geisser correction was conducted to determine the effects of nudges ('no nudge' and 'nudge (D)') and load ('NL', 'LL' and 'HL') on RTs of the audit task **(Panel 2B)**. The main effects of nudges and load were both significant, F (1, 144) = 466.89, $p < 0.001$, η2 = 0.76, F (1.98, 288) = 11.69, $p < 0.001$, η2 = 0.075. The nudges x load interaction was also highly significant, F (1.98, 288) = 47.64, $p < 0.001$, η2 = 0.25 **(Panel 3A)**. Similarly, for the audit task accuracy **(Panel 2E)**, the main effects of nudges and load were significant, F (1, 144) = 9.38, $p = 0.003$, η2 = 0.06, F (1.93, 288) = 1049, $p < 0.001$, η2 = 0.07. Likewise, the nudges x load interaction was significant, F (1.93, 288) = 10.5, $p < 0.001$, η2 = 0.068 **(Panel 3B)**.

Post-hoc analysis with Bonferroni correction was conducted for pairwise comparisons. For RTs in NL, LL, and HL conditions, no nudge vs. nudge (D) groups were significantly different (MD = 0.076, 95% CI, 0.03 to 0.11, p< 0.001), (MD = 0.25, 95% CI, 0.22 to 0.29, p< 0.001) and (MD = 0.34, 95% CI, 0.31 to 0.38, p< 0.001), respectively. In the no nudge group, NL vs. LL (MD = -0.13, 95% CI, -0.17 to -0.09, p< 0.001), NL vs. HL (MD = -0.19, 95% CI, -0.23 to -0.16, p<

0.001), and LL vs. HL (MD = -0.06, 95% CI, -0.09 to -0.03, *p*< 0.001) were all significantly different. In the nudge group while NL vs. LL (MD = 0.04, 95% CI, -0.08 to 0.1, *p* = 0.28), and LL vs. HL (MD = 0.03, 95% CI, -0.03 to -0.08, *p* = 0.78), were not significant, NL vs. HL (MD = 0.07, 95% CI, 0.01 to 0.13, *p* = 0.01) was significant (9 pairwise comparisons).

Subsequently, post-hoc pairwise comparison for accuracy of audit task across load conditions showed no difference between no nudge and nudge (D) (MD = 2.08, 95% CI, -3.84 to 8.1, *p* = 0.49) in NL whereas in LL (MD = -10.88, 95% CI, -17.48 to -4.27, *p* = 0.001) and HL (MD = -13.16, 95% CI, -19.37 to -6.95, *p* < 0.001), there was significant differences between the groups. In no nudge group, NL vs. LL (MD = -0.93, 95% CI, -5.62 to 3.76, *p* = 0.99), NL vs. HL (MD = 0.83, 95% CI, -3.76 to 5.42, *p* = 0.99), and LL vs. HL (MD = 1.76, 95% CI, -2.23 to 5.75, *p* = 0.86) were all non-significant. In the nudge group while NL vs. LL (MD = -13.89, 95% CI, -21.8 to -5.98, *p* < 0.001), and NL vs. HL (MD = -14.41, 95% CI, -22.16 to -6.67, *p* < 0.001), were significant, LL vs. HL (MD = -0.53, 95% CI, -7.25 to 6.2, *p* = 0.99) was not (9 pairwise comparisons).

Overall, the presence of default nudges facilitates response efficiency by reducing RTs and enhancing accuracy. In contrast, increasing load imposes a cognitive burden, leading to prolonged RTs while exerting no significant effect on accuracy. However, the interaction between these factors suggests that nudges moderate the impact of load, attenuating the RT cost associated with higher load levels while concurrently improving accuracy.

Experiment 3b: The effectiveness of default nudges pertains to their contextual placement

Results

The Wilcoxon signed-rank test 'nudge (A) LL' and 'nudge (A) HL' with within-subject factor on RTs showed Z = -4.533, *p* < 0.001 **(Panel 2C)** and on accuracy rate Z = 5.089, *p* < 0.001 **(Panel 2F)**, evidencing the cognitive load was also successfully modulated in the nudge (A) group, replicating the experiment 1 and 2.

A two-way mixed ANOVA with Greenhouse-Geisser correction was done to examine the effects of the between-subject factor of 'no nudge', 'nudge (D)' and 'nudge (A)' and load with the within-subject factor of 'NL', 'LL' and 'HL' on RTs of the audit task. The analysis revealed a significant main effect of nudges on RTs (F (2, 201) = 392.65, *p* < 0.001, η2 = 0.77) **(Panel 2C)**. On the other hand, the main effect of the load was not significant (F (1.87, 402) = 1.31, *p* = 0.27, η2 = 0.006). The interaction of nudges x load was significant (F (3.74, 402) = 52.36, *p* < 0.001, η2 = 0.34) **(Panel 3A)**. The analysis further showed that the main effect of nudges on accuracy was significant (F (2, 201) = 84.26, *p* < 0.001, η2 = 0.46) **(Panel 2F)**. However, the main effect of load was not significant (F (1.84, 402) = 0.423, *p* = 0.63, η2 = 0.002). Furthermore, the nudges x load interaction was significant (F (3.69, 402) = 22.9, *p* < 0.001, η2 = 0.19) **(Panel 3B)**. These results highlight that the effect of nudges depended on the presence of load. Thus, no post-hoc analyses were conducted for the main effect of load on either RTs or accuracy.

For RTs at NL, no nudge vs. nudge (A), and no nudge vs. nudge (D) were significant (MD = 0.068, 95% CI, 0.03 to 0.11, *p* < 0.001) and (MD = 0.08, 95% CI, 0.03 to 0.12, *p* < 0.001), and nudge (A) vs. nudge (D) were not significant (MD = 0.008, 95% CI, -0.04 to 0.061, *p* = 0.99). At LL, while no nudge vs. nudge (A) and no nudge vs. nudge (D) were significant (MD = 0.25, 95% CI, 0.21 to 0.29, *p* < 0.001) and (MD = 0.25, 95% CI, 0.21 to 0.3, *p* < 0.001), nudge (A) vs. nudge (D) were not significant (MD = 0.006, 95% CI, -0.04 to 0.06, *p* = 0.99). Finally at HL, no nudge vs. nudge (A) and no nudge vs. nudge (D) were significant (MD = 0.36, 95% CI, 0.32 to 0.39, *p*

< 0.001) and (MD = 0.34, 95% CI, 0.3 to 0.38, $p < 0.001$), whereas nudge (A) vs. nudge (D) were not significant (MD = -0.01, 95% CI, -0.059 to 0.032, $p = 0.99$) (9 pairwise comparisons).

For accuracy, at NL, only no nudge vs. nudge (A) was significant (MD = 8.85, 95% CI, 2.25 to 15.44, $p = 0.004$), but no nudge vs. nudge (D) and nudge (A) vs. nudge (D) were not significant (MD = 2.089, 95% CI, -5.55 to 9.73, $p = 0.99$) (MD = -6.76, 95% CI, -15.22 to 1.69, $p = 0.16$). At LL, no nudge vs. nudge (A), no nudge vs. nudge (D) and nudge (A) vs. nudge (D) were all significant (MD = 26.78, 95% CI, 20.51 to 33.05, $p < 0.001$), (MD = -10.88, 95% CI, -18.14 to -3.61, $p = 0.001$), and (MD = -37.66, 95% CI, -45.69 to -29.62, $p < 0.001$). At HL, all comparisons were also significant including no nudge vs. nudge (A) (MD = 25.54, 95% CI, 19.88 to 31.2, $p < 0.001$), no nudge vs. nudge (D) (MD = -13.16, 95% CI, -19.72 to -6.61, $p < 0.001$), nudge (A) vs. nudge (D) (MD = -38.7, 95% CI, -45.96 to -31.45, $p < 0.001$) (9 pairwise comparisons).

These results suggest that, regardless of its placement, the default nudges reduce RTs, enhancing the speed of performance in the audit task. However, its effectiveness depends on the presence of the load and the specific placement of the nudges. Under load, nudges accommodate accuracy when applied to the response option with a higher probability of correctness, it can reduce accuracy when placed on the response with a lower probability.

Default nudges placement shifts the balance between detecting aggressive items and avoiding false alarm

In order to determine the effect of nudges on objective scepticism across cognitive load levels, a series of Kruskal-Wallis tests were conducted which reveal significant effects in all conditions: χ^2 (2) = 40.41, $p < 0.001$ (NL), χ^2 (2) = 52.67, $p < 0.001$ (LL), and χ^2 (2) = 44.49, $p < 0.001$ (HL). To further investigate these effects, Mann-Whitney U tests were performed for pairwise comparisons with adjusted *p*-values. In the NL condition, significant differences were observed between no nudge and nudge (D) ($p = 0.01$), no nudge and nudge (A) ($p < 0.001$), and nudge (D) and nudge (A) ($p < 0.001$). In the LL condition, significant differences emerged between no nudge and nudge (D) ($p < 0.001$), no nudge and nudge (A) ($p < 0.001$), and nudge (D) and nudge (A) ($p < 0.001$). In the HL condition, performance accuracy also significantly differed between no nudge and nudge (D) ($p < 0.001$), no nudge and nudge (A) ($p < 0.001$), and nudge (D) and nudge (A) ($p < 0.001$). Putting together, while positioning the default nudges on responses with high probability of correctness enhances auditor overall performance (as shown in the previous sections), placing the default nudges on low-probability responses appears to promote greater scepticism **(Figure 3, Panel A).**

To further explore the impact of default nudges placement on the correct detection of aggressive and non-aggressive items, we analysed the general response trends across all three experimental groups ('no nudge', 'nudge (A)', and 'nudge (D)') **(Figure 3)**. The Kruskal-Wallis test revealed significant differences across cognitive load conditions for all response types: true positives (TP: $\chi^2(2) = 138.262$, $p < 0.001$), false positives (FP: $\chi^2(2) = 113.559$, $p < 0.001$), false negatives (FN: $\chi^2(2) = 138.1$, $p < 0.001$), and true negatives (TN: $\chi^2(2) = 112.771$, $p = 0.001$). Pairwise comparisons using the Mann-Whitney U test, with Bonferroni-adjusted *p*-values, indicated that TP rates differed significantly between no nudge and nudge (D) ($p < 0.001$), no nudge and nudge (A) ($p < 0.001$), and nudge (D) and nudge (A) ($p < 0.001$). A similar pattern emerged for FP, FN, and TN rates, with all pairwise comparisons yielding significant adjusted *p*-values ($p < 0.001$). These results suggest that nudging toward a response with low probabilty of correctness boosts auditors' ability to accurately detect aggressive items (Hit, **Panel B**). However, it increases the likelihood of failing to detect an aggressive item, leading to its misclassification as non-aggressive (Miss, **Panel C**). When the nudges are placed on the response with more probable correctness, it whereas auditors may exhibit an elevated tendency to misclassify non-aggressive items as aggressive (False alarm, **Panel D**). Yet, it can help the correct identification of non-aggressive items (Correct rejection, **Panel E**).

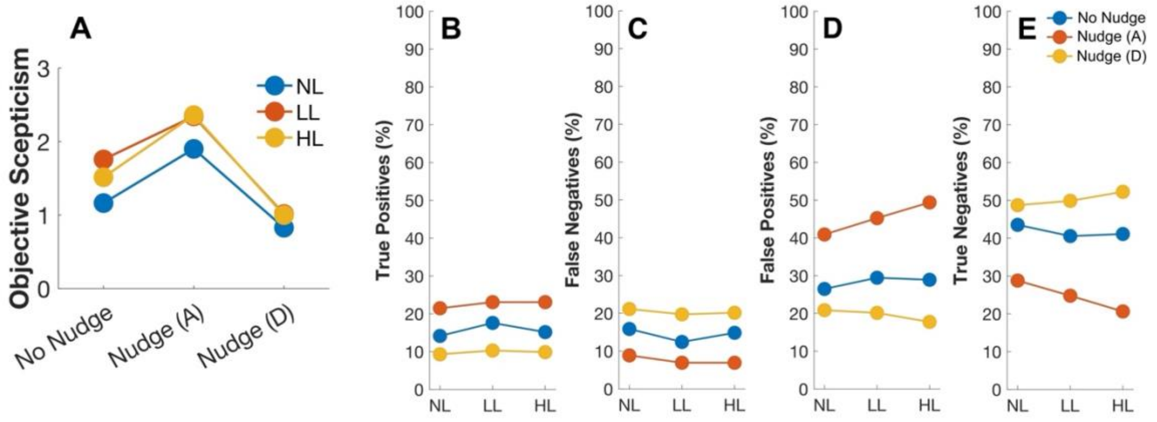


Figure 3. Panel ***A*** *shows objective scepticism across the no nudge, nudge (A), and nudge (D) groups under NL, LL, and HL conditions. Panels* ***B***, ***C***, ***D***, *and* ***E*** *display the objective scepticism across nudges conditions for each outcome type: true positives (Panel* ***B****), false negatives (Panel* ***C****), false positives (Panel* ***D****), and true negatives (Panel* ***E****).*

Discussion

In an online study, we investigated how default nudges influence professional scepticism under varying levels of cognitive load among auditors. Experiment 1 assessed task-induced cognitive load (minimal, low, medium, high) using accuracy and RTs to establish behavioural dissociability. Results indicated that minimal, low, and medium load conditions differed significantly from the high load condition in terms of accuracy, while RTs increased in proportion to the level of cognitive load. Based on these findings, the low load and high load conditions were selected for subsequent experiments.

In experiment 2, we examined the impact of cognitive load (NL, LL, HL) on audit task performance by imposing short-term memory demands. Performance was assessed in terms of RTs and accuracy. We found that RTs increased proportionally with cognitive load in the audit task. However, accuracy remained unaffected across conditions.

In experiment 3a, we examined how default nudges affect audit task performance across three levels of cognitive load, comparing a no-nudge group with a nudge group. Default nudges effectively accelerated RTs, but their impact on accuracy was load-dependent, emerging only under elevated cognitive demands. In experiment 3b, we tested whether the effectiveness of nudges depended on their alignment with the response option associated with higher probability of correctness. Nudges had no impact under NL conditions. However, under cognitive load, default nudges led to shorter RTs and had a significant effect on accuracy. Under both LL and HL, accuracy increased when the nudge was applied to the 'Disagree' option, the response with a higher probability of correctness, and decreased when placed on 'Agree'.

Cognitive load modulates both the speed and accuracy of the WM task

Our findings show that modulating cognitive load significantly influences RTs and impedes accuracy, both of which are key indicators of WM function. Elevated cognitive load resulted in prolonged RTs and reduced accuracy, reflecting load-dependent constraints on WM capacity, consistent with the notion that increased intrinsic load or task complexity depletes available WM resources (Sweller, 1988). These results are in line with previous evidence showing that higher memory load disrupts accuracy and slows responses, highlighting inherent fundamental capacity limitations within cognitive processing (Barrouillet et al., 2004; Barrouillet et al., 2007; Camos & Portrat, 2015; Leung et al., 2004; Simon & Wolf, 1963; Zhang et al., 2011; Zhao et al., 2010).

In spatial WM tasks, excessive load impairs both the maintenance, the ability to hold spatial location information in mind over time, and the manipulation, which involves updating or rearranging these locations to make decisions or solve problems, of task-relevant information (Leung et al., 2004). Similarly, in response selection paradigms, such as the Simon effect, increasing verbal WM load abolishes the facilitation typically observed for congruent responses, underscoring WM role in conflict resolution (Zhao et al., 2010). The Simon effect refers to the phenomenon where people respond faster and more accurately when the spatial location of a stimulus matches the location of the required response (congruent), even when the stimulus location is irrelevant to the task (Simon & Wolf, 1963). Under low cognitive load, participants exhibit faster responses when stimulus and response locations align, however, this advantage diminishes with rising load, suggesting that higher cognitive demands impair efficient response selection. Furthermore, cognitive load modulates attentional control, with lower loads permitting more effective attentional allocation, while higher loads disrupt attentional guidance (Zhang et al., 2011). Studies on delayed recall further support this load-dependent interference, as increased concurrent task demands impair both immediate and delayed recall, highlighting the vulnerability of maintenance and retrieval processes to heightened cognitive demands (Camos & Portrat, 2015).

Cognitive load adversely influences the speed but not the accuracy of DM-based professional scepticism

The results of the audit task indicate a decrease in RTs proportional to the increase in cognitive load. This slowdown is likely attributable to the additional cognitive demands required to maintain and process information under higher load conditions (Barrouillet et al., 2007; Lam et al., 2022; Langerock et al., 2025; Miri Ashtiani & Daliri, 2023; Pei et al., 2023; Sisakhti et al., 2021). Despite the observed reduction in professional scepticism task speed, accuracy remains unaffected, suggesting that the preservation of accuracy may reflect a compensatory strategy. Specifically, participants may adopt more deliberate DM-based professional scepticism processes in response to cognitive load. This, in turn, may lead to slower, yet accurate responses. This finding supports the *speed-accuracy trade-off* effect, which posits that in tasks where accuracy is critical, such as audit tasks, individuals may intentionally slow their DM to guarantee correctness, even at the cost of speed, in order to minimize errors (Donkin et al., 2014; Forstmann et al., 2008; Garrett, 1922; Hick, 1952; Ollman, 1966; Pachella, 2021; Ratcliff & Rouder, 1998; Schouten & Bekker, 1967; Wickelgren, 1977; Woodworth, 1899). In our audit task, participants were instructed to respond at their 'own pace', with no time constraints, which may explain why increased cognitive load did not affect accuracy. Nonetheless, under conditions where both load and time constraints are high, individuals may be compelled to make quicker decisions, often at the expense of accuracy (Edwards et al., 2012; Onagawa et al., 2019; Rieskamp & Hoffrage, 2008).

Like with our results, Becker and colleagues (Becker et al., 2016) proposed that in tasks with low cognitive load (associated with our NL), responses tend to be automated, with slower RTs often indicating lapses in attention which leads to negative correlations between RTs and accuracy (e.g., faster responses are likely to be correct). In contrast, tasks with high cognitive load (corresponding with our LL, and HL) necessitate controlled and sequential processing, resulting in positive correlations between RTs and accuracy (e.g., slower responses are more likely to be correct). Additionally, given that RTs varied proportionally to the load level, dual-process theory provides a useful framework for understanding these findings. According to this theory, Type 2 processes, which are rule-based, analytical, *slow*, controlled, and conscious, are associated with higher success rates and fewer errors, in contrast to Type 1 processes, which are *fast*, automatic, and non-conscious (Frankish, 2010). Empirical evidence indicates that Type 2 processing is particularly susceptible to the impairing effects of cognitive load such as time pressure (Harris & Santhosh, 2022).

Default nudges promote DM-based professional scepticism under cognitive load provided that it is positioned correctly

We observed that the implementation of default nudges did not lead to improvement in professional scepticism performance, neither in response speed nor accuracy, when comparing the no nudge and nudge conditions in the absence of cognitive load. Since individuals were not constrained by cognitive load, they did not blindly follow the nudges. Instead, having more time and cognitive resources allowed them to analyse audit pieces more deliberately, even if it led to conscious errors. This supports the view that default nudges primarily function by reducing the *cognitive load* associated with complex tasks (Meunier et al., 2024). In addition, Van Gestel and colleagues (Van Gestel et al., 2021) state that when individuals have the opportunity to engage in effortful deliberation, they can override default nudges through extensive reasoning. This implies that default effects are robust, but not necessarily impregnable as deliberation may operate in parallel and impact choice simultaneously.

However, under cognitive load, a strikingly different pattern emerged. When the default nudges aligned with the responses with high probability of correctness, decision accuracy improved relative to the no-nudge condition. But when the nudges were placed on the less probable response, accuracy significantly declined. There are two mutually inclusive possibilities for this. The first possibility suggests that individuals frequently shift processing strategies under cognitive load, often relying on simple heuristics, which can preserve or even improve decision accuracy in certain contexts. This adaptive strategy may help explain the success observed in audit task performance (Brunye et al., 2018; Gigerenzer & Gaissmaier, 2011; Sisakhti et al., 2021). This suggests that under cognitive load, individuals show a strong propensity to adhere to the default option, regardless of its alignment with the optimal choice. Further, this highlights the critical role of default nudges architecture in ensuring its effectiveness, particularly when cognitive resources are taxed. The second possibility is that under Type 1 processing, nudges become more effective no matter where they lead individuals to or even, guarantee success. That is, as nudges take advantage of the heuristics and biases that are correlated with Type 1 processing, they are more effective in exactly those circumstances where people are most likely to use Type 1 processing. Therefore, inhibiting Type 2 reasoning, for example, by time pressure (Evans & Curtis-Holmes, 2005) or concurrent WM load through more information (Bago & De Neys, 2017) could in theory increase the effectiveness of nudges.

The presence of nudges did not alter the performance in LL relative to HL. These findings converge with those of (Van Gestel et al., 2021), who demonstrated that the efficacy of the default nudges remains invariant across cognitive load conditions (LL vs. HL). In their study, despite participants experiencing greater difficulty in recalling the dot pattern under HL (demonstrated through RTs), the default effect remained statistically equivalent. Notably, their study did not incorporate an NL condition. They proposed that default nudges are most effective when Type 2 processing is inhibited. This conforms with the default-interventionist model of dual processing (Evans & Stanovich, 2013), which contends that Type 1 processing constitutes the baseline mode of cognition, operating autonomously unless actively overridden by Type 2 reasoning. Crucially, the mere availability of cognitive resources (e.g., in LL condition) does not necessitate the individuals' higher engagement in the task at hand (Van Gestel et al., 2021).

Strategic placement of default nudges can serve different purposes

Our findings demonstrate that the placement of default nudges has significant and nuanced effects on DM-based professional scepticism in high-stakes auditing contexts. Nudges corresponding to the responses with higher probability of correctness reliably enhanced overall professional scepticism performance across all cognitive load conditions (e., LL and HL). Specifically, they were associated with higher overall accuracy and faster RTs, suggesting that these defaults leveraged intuitive, Type 1 processing to streamline judgment under both low and high cognitive demands (Evans & Stanovich, 2013). This efficiency gain is particularly valuable in complex environments where information overload can compromise deliberative reasoning (Gigerenzer & Gaissmaier, 2011; Payne et al., 1993).

However, the placement of defaults on such responses is not without trade-offs. While it increased true negative rates and reduced false positives, thereby improving accuracy in non-aggressive classifications, it also increased false negatives, e.g., instances in which aggressive audit items were missed. This pattern reveals a conservative bias induced by the default: auditors become more likely to follow the expected norm, potentially overlooking critical red flags (Jachimowicz et al., 2019). In practice, this could translate into missed detections of material misstatements or fraud, posing legal and reputational risks for firms. Therefore, while such nudges configuration optimizes average performance, it may inadvertently downplay low-probability but high-cost errors.

In contrast, placing the default nudge on the less probable response induced a more sceptical decision pattern. This configuration significantly increased the detection of aggressive items (true positives) and reduced false negatives, suggesting that auditors became more vigilant in identifying problematic cases. Importantly, this vigilance came at the cost of increased false positives, and non-aggressive items incorrectly flagged as aggressive, which could contribute to unnecessary escalation, audit inefficiencies, or strained client relationships. Yet from a risk management perspective, such over-detection may be desirable in contexts where the cost of a false negative (e.g., missing a fraud indicator) far outweighs the cost of a false positive (e.g., reviewing a benign case). These findings are consistent with research suggesting that when error costs are asymmetrical, nudging individuals toward the less likely, but more cautious, response can serve as a protective mechanism against catastrophic failures (Arkes et al., 1986; Gigerenzer & Edwards, 2003).

Thus, the decision to place a default on the responses with high or low probability of correctness should be informed by the specific error-cost asymmetries of the task domain. In domains like auditing and medicine, where the consequences of false negatives can be severe, legal liability, patient harm, regulatory sanction, and defaults on less probable, risk-sensitive options may serve as a safeguard. Conversely, in settings where efficiency and throughput are critical, and the cost of false positives is more salient, defaults aligned with the more probable response may optimize decision fluency and minimize unnecessary friction. Crucially, our findings suggest that default nudges are not neutral tools; their effectiveness and ethical soundness depend on how well their configuration aligns with domain-specific risk profiles and human cognitive constraints (Reijula & Hertwig, 2022; Thaler & Sunstein, 2008).

Conclusion

This study highlights the complex relationship between default nudges and DM-based professional scepticism under cognitive load. Our findings reveal that cognitive load impairs professional scepticism, potentially compromising auditors’ ability to detect fraud and assess risks accurately. However, default nudges can mitigate these effects, accelerating RTs and improving audit accuracy, but only when cognitive load is present. The effectiveness of nudges depends on their alignment with pre-existing response tendencies, with scepticism-based accuracy benefiting when nudges reinforce the response with high probability of correctness. These results highlight the potential for carefully designed choice architectures to support DM in high-stakes environments, ensuring that cognitive constraints do not undermine critical judgment. Future research should explore the long-term implications of nudges interventions and their application in real-world auditing contexts.

References

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

Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational behavior and human decision processes*, *37*(1), 93-110.

Asare, S. K., Trompeter, G. M., & Wright, A. M. (2000). The effect of accountability and time budgets on auditors' testing strategies. *Contemporary Accounting Research*, *17*(4), 539-560. https://doi.org/10.1506/F1EG-9EJG-DJ0B-JD32

Austin, A. A. (2023). Remembering fraud in the future: Investigating and improving auditors' attention to fraud during audit testing. *Contemporary Accounting Research*, *40*(2), 925-951. https://doi.org/10.1111/1911-3846.12843

Backs, R. W., & Seljos, K. A. (1994). Metabolic and cardiorespiratory measures of mental effort: the effects of level of difficulty in a working memory task. *International Journal of psychophysiology*, *16*(1), 57-68. https://doi.org/10.1016/0167-8760(94)90042-6

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, *158*, 90-109. https://doi.org/10.1016/j.cognition.2016.10.014

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of experimental psychology: General*, *133*(1), 83.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *J Exp Psychol Learn Mem Cogn*, *33*(3), 570-585. https://doi.org/10.1037/0278-7393.33.3.570

Barrouillet, P., & Camos, V. (2012). As time goes by: Temporal constraints in working memory. *Current directions in psychological science*, *21*(6), 413-419. https://doi.org/10.1177/0963721412459513

Barzykowski, K., & Niedzwienska, A. (2018). Involuntary autobiographical memories are relatively more often reported during high cognitive load tasks. *Acta Psychol (Amst)*, *182*, 119-128. https://doi.org/10.1016/j.actpsy.2017.11.014

Baumeister, R. F. (2001). Ego depletion, the executive function, and self-control: An energy model of the self in personality. https://doi.org/10.1080/152988602317319302

Becker, N., Schmitz, F., Göritz, A. S., & Spinath, F. M. (2016). Sometimes more is better, and sometimes less is better: Task complexity moderates the response time accuracy correlation. *Journal of Intelligence*, *4*(3), 11. https://doi.org/10.3390/jintelligence4030011

Bedeir, R. E. (2024). The differential impact of distracted auditors in managing portfolio of financially distressed audit clients on audit quality: the role of professional skepticism. *Future Business Journal*, *10*(1), 36. https://doi.org/10.1186/s43093-024-00321-9

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 12. https://doi.org/10.1037/0096-1523.14.1.12

Bianco, R., Mills, G., de Kerangal, M., Rosen, S., & Chait, M. (2021). Reward Enhances Online Participants' Engagement With a Demanding Auditory Task. *Trends Hear*, *25*, 23312165211025941. https://doi.org/10.1177/23312165211025941

Board, P. C. A. O. (2012). Maintaining and Applying Professional Skepticism in Audits. Staff Audit Practice Alert No. 10. In: PCAOB Washington, DC.

Bohn, R., & Short, J. E. (2012). Info capacity| measuring consumer information. *International Journal of Communication*, *6*, 21.

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, *402*(6758), 179-181. https://doi.org/10.1038/46035

Botvinick, M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cogn Affect Behav Neurosci*, *7*(4), 356-366. https://doi.org/10.3758/cabn.7.4.356

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

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

Brunye, T. T., Martis, S. B., & Taylor, H. A. (2018). Cognitive load during route selection increases reliance on spatial heuristics. *Q J Exp Psychol (Hove)*, *71*(5), 1045-1056. https://doi.org/10.1080/17470218.2017.1310268

Camos, V., & Portrat, S. (2015). The impact of cognitive load on delayed recall. *Psychon Bull Rev*, *22*(4), 1029-1034. https://doi.org/10.3758/s13423-014-0772-5

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*(5364), 747-749. https://doi.org/10.1126/science.280.5364.747

Cohen, J. (1988). Statistical power analysts for the behavioral sciences. 2nd edn Hillsdale. *NJL Erlbaum Associates*.

Csipo, T., Lipecz, A., Mukli, P., Bahadli, D., Abdulhussein, O., Owens, C. D., Tarantini, S., Hand, R. A., Yabluchanska, V., & Kellawan, J. M. (2021). Increased cognitive workload evokes greater neurovascular coupling responses in healthy young adults. *PLoS One*, *16*(5), e0250043. https://doi.org/10.1371/journal.pone.0250043

Donkin, C., Little, D. R., & Houpt, J. W. (2014). Assessing the speed--accuracy trade-off effect on the capacity of information processing. *J Exp Psychol Hum Percept Perform*, *40*(3), 1183-1202. https://doi.org/10.1037/a0035947

Edwards, J. A., Snyder, F. J., Allen, P. M., Makinson, K. A., & Hamby, D. M. (2012). Decision making for risk management: a comparison of graphical methods for presenting quantitative uncertainty. *Risk Anal*, *32*(12), 2055-2070. https://doi.org/10.1111/j.1539-6924.2012.01839.x

Evans, J. S., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspect Psychol Sci*, *8*(3), 223-241. https://doi.org/10.1177/1745691612460685

Evans, J. S. B., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, *11*(4), 382-389. https://doi.org/10.1080/13546780542000005

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E. J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proc Natl Acad Sci U S A*, *105*(45), 17538-17542. https://doi.org/10.1073/pnas.0805903105

Frankish, K. (2010). Dual-process and dual-system theories of reasoning. *Philosophy Compass*, *5*(10), 914-926. https://doi.org/10.1111/j.1747-9991.2010.00330.x

Gailliot, M. T. (2008). Unlocking the Energy Dynamics of Executive Functioning: Linking Executive Functioning to Brain Glycogen. *Perspect Psychol Sci*, *3*(4), 245-263. https://doi.org/10.1111/j.1745-6924.2008.00077.x

Gajewski, Heimann, M., Léger, P.-M., & Teye, P. (2024). Enhancing auditors' professional skepticism through nudges: an eye-tracking experiment. *Accounting and Business Research*, 1-19. https://doi.org/10.1080/00014788.2024.2364215

Gajewski, Heimann, M., Meunier, L., & Ohadi, S. (2024). Nudges for responsible finance? A survey of interventions targeted at financial decision making. *Corporate Social Responsibility and Environmental Management*, *31*(2), 1203-1219. https://doi.org/10.1002/csr.2625

Gajewski, J.-F., Heimann, M., & Meunier, L. (2022). Nudges in SRI: the power of the default option. *Journal of Business Ethics*, 1-20. https://doi.org/10.1007/s10551-020-04731-x

Garrett, H. E. (1922). *A study of the relation of accuracy to speed* (Vol. 8). Columbia university.

Gigerenzer, G. (2008). Why Heuristics Work. *Perspect Psychol Sci*, *3*(1), 20-29. https://doi.org/10.1111/j.1745-6916.2008.00058.x

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: from innumeracy to insight. *Bmj*, *327*(7417), 741-744. https://doi.org/10.1136/bmj.327.7417.741

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, *62*(1), 451-482. https://doi.org/10.1146/annurev-psych-120709-145346

González-Díaz, B., & García-Fernández, R. (2018). Auditing. *Global Encyclopedia of Public Administration, Public Policy, and Governance*, 328-345. https://doi.org/https://doi.org/10.1007/978-3-319-20928-9_2362

Harris, E., & Santhosh, L. (2022). Dual Process Theory and Cognitive Load: How Intensivists Make Diagnoses. *Critical Care Clinics*, *38*(1), 27-36. https://doi.org/10.1016/j.ccc.2021.07.001

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of experimental psychology*, *4*(1), 11-26. https://doi.org/10.1080/17470215208416600

Hu, J., Xin, C., Zhang, M., & Chen, Y. (2024). The effect of cognitive load and time stress on prospective memory and its components. *Current Psychology*, *43*(2), 1670-1684. https://doi.org/10.1007/s12144-023-04354-1

Hurtt, R. K. (2010). Development of a scale to measure professional skepticism. *Auditing: A Journal of Practice & Theory*, *29*(1), 149-171. https://doi.org/10.2308/aud.2010.29.1.149

Hurtt, R. K., Brown-Liburd, H., Earley, C. E., & Krishnamoorthy, G. (2013). Research on auditor professional skepticism: Literature synthesis and opportunities for future research. *Auditing: A Journal of Practice & Theory*, *32*(Supplement 1), 45-97. https://doi.org/10.2308/ajpt-50361

Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, *3*(2), 159-186. https://doi.org/10.1017/bpp.2018.43

Johari, R. J., Ridzoan, N. S., & Zarefar, A. (2019). The influence of work overload, time pressure and social influence pressure on auditors' job performance. *International Journal of Financial Research*, *10*(3), 88-106. https://doi.org/10.5430/ijfr.v10n3p88

Kashay, A. Q., Cheung, J. Y., Vaknalli, R. N., Delaney, M. J., Navarro Jr, M. B., Junaidi, C., Veenker, F., Neuwirth, M. E., Gabriel, C. J., & DeNardo, L. A. (2022). Neural activity in the anterior cingulate cortex is required for effort-based decision making. *BioRxiv*, 2022.2003. 2022.485350. https://doi.org/10.1101/2022.03.22.485350

Kerns, J. G., Cohen, J. D., MacDonald, A. W., 3rd, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*(5660), 1023-1026. https://doi.org/10.1126/science.1089910

Kool, W., & Botvinick, M. (2018). Mental labour. *Nature human behaviour*, *2*(12), 899-908. https://doi.org/10.1038/s41562-018-0401-9

Kunda, Z., & Oleson, K. C. (1995). Maintaining stereotypes in the face of disconfirmation: constructing grounds for subtyping deviants. *Journal of personality and social psychology*, *68*(4), 565. https://doi.org/10.1037//0022-3514.68.4.565

Lam, T. K., Vartanian, O., & Hollands, J. G. (2022). The brain under cognitive workload: Neural networks underlying multitasking performance in the multi-attribute task battery. *Neuropsychologia*, *174*, 108350. https://doi.org/10.1016/j.neuropsychologia.2022.108350

Lambert, T. A., Jones, K. L., Brazel, J. F., & Showalter, D. S. (2017). Audit time pressure and earnings quality: An examination of accelerated filings. *Accounting, Organizations and Society*, *58*, 50-66. https://doi.org/10.1016/j.aos.2017.03.003

Langerock, N., Oberauer, K., Throm, E., & Vergauwe, E. (2025). The cognitive load effect in working memory: Refreshing the empirical landscape, removing outdated explanations. *Journal of Memory and Language*, *140*, 104558. https://doi.org/10.1016/j.jml.2024.104558

Leung, H. C., Seelig, D., & Gore, J. C. (2004). The effect of memory load on cortical activity in the spatial working memory circuit. *Cogn Affect Behav Neurosci*, *4*(4), 553-563. https://doi.org/10.3758/cabn.4.4.553

Li, H. L., & van Rossum, M. C. (2020). Energy efficient synaptic plasticity. *Elife*, *9*. https://doi.org/10.7554/eLife.50804

Lopez, D. M., & Greenwald, S. M. (2022). A taxing audit—On the association between auditor workload compression and tax avoidance. *International Journal of Auditing*, *26*(4), 420-445. https://doi.org/10.1111/ijau.12289

Mansouri, Koechlin, E., Rosa, M. G. P., & Buckley, M. J. (2017). Managing competing goals - a key role for the frontopolar cortex. *Nat Rev Neurosci*, *18*(11), 645-657. https://doi.org/10.1038/nrn.2017.111

Mazzoni, G. (2019). Involuntary memories and involuntary future thinking differently tax cognitive resources. *Psychol Res*, *83*(4), 684-697. https://doi.org/10.1007/s00426-018-1123-3

McDaniel, L. S. (1990). The effects of time pressure and audit program structure on audit performance. *Journal of accounting research*, *28*(2), 267-285. https://doi.org/10.2307/2491150

Mertens, S., Herberz, M., Hahnel, U. J., & Brosch, T. (2022). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proc. Natl. Acad. Sci.* https://doi.org/doi.org/10.1073/pnas.2107346118

Meunier, L., Bashirzadeh, Y., & Ohadi, S. (2024). Framing the Default Option Right. *Journal of Behavioral Decision Making*, *37*(3), e2395. https://doi.org/10.1002/bdm.2395

Meunier, L., & Richit, S. (2024). Testing four nudges in socially responsible investments: Default winner by inertia. *Business ethics, the environment & responsibility*, *33*(3), 392-415. https://doi.org/10.1111/beer.12612

Miller, J. (1991). Reaction time analysis with outlier exclusion: bias varies with sample size. *Q J Exp Psychol A*, *43*(4), 907-912. https://doi.org/10.1080/14640749108400962

Miller, J. (2023). Outlier exclusion procedures for reaction time analysis: The cures are generally worse than the disease. *J Exp Psychol Gen*, *152*(11), 3189-3217. https://doi.org/10.1037/xge0001450

Miri Ashtiani, S. N., & Daliri, M. R. (2023). Identification of cognitive load-dependent activation patterns using working memory task-based fMRI at various levels of difficulty. *Sci Rep*, *13*(1), 16476. https://doi.org/10.1038/s41598-023-43837-w

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *J Exp Psychol Gen*, *130*(4), 621-640. https://doi.org/10.1037//0096-3445.130.4.621

Münscher, R., Vetter, M., & Scheuerle, T. (2016). A review and taxonomy of choice architecture techniques. *Journal of Behavioral Decision Making*, *29*(5), 511-524. https://doi.org/10.1002/bdm.1897

Nelson, M. W. (2009). A model and literature review of professional skepticism in auditing. *Auditing: A Journal of Practice & Theory*, *28*(2), 1-34. https://doi.org/10.2308/aud.2009.28.2.1

Ninomiya, Y., Iwata, T., Terai, H., & Miwa, K. (2024). Effect of cognitive load and working memory capacity on the efficiency of discovering better alternatives: A survival analysis. *Mem Cognit*, *52*(1), 115-131. https://doi.org/10.3758/s13421-023-01448-w

Normand, M. P. (2008). Science, skepticism, and applied behavior analysis. *Behav Anal Pract*, *1*(2), 42-49. https://doi.org/10.1007/BF03391727

Ollman, R. (1966). Fast guesses in choice reaction time. *Psychonomic Science*, *6*(4), 155-156.

Onagawa, R., Shinya, M., Ota, K., & Kudo, K. (2019). Risk aversion in the adjustment of speed-accuracy tradeoff depending on time constraints. *Sci Rep*, *9*(1), 11732. https://doi.org/10.1038/s41598-019-48052-0

Ordali, E., Marcos-Prieto, P., Avvenuti, G., Ricciardi, E., Boncinelli, L., Pietrini, P., Bernardi, G., & Bilancini, E. (2024). Prolonged exertion of self-control causes increased sleep-like frontal brain activity and changes in aggressivity and punishment. *Proc Natl Acad Sci U S A*, *121*(47), e2404213121. https://doi.org/10.1073/pnas.2404213121

Otley, D. T., & Pierce, B. J. (1996). The operation of control systems in large audit firms. *Auditing*, *15*(2), 65.

Pachella, R. G. (2021). The interpretation of reaction time in information-processing research 1. In *Human information processing* (pp. 41-82). Routledge.

Padamsey, Z., & Rochefort, N. L. (2023). Paying the brain's energy bill. *Curr Opin Neurobiol*, *78*, 102668. https://doi.org/10.1016/j.conb.2022.102668

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge university press.

PCAOB. (2022). Staff Update and Preview of 2021 Inspection Observations. https://pcaob-assets.azureedge.net/pcaob-dev/docs/default-source/documents/staff-preview-2021-inspection-observations-spotlight.pdf?sfvrsn=d2590627_2/

PCAOB. (2023). *Professional Competence and Skepticism are Essential to Quality Audits*. Retrieved from Available from: https://pcaobus.org/documents/competence-and-skepticism-spotlight.pdf(open in a new window)

Pei, L., Northoff, G., & Ouyang, G. (2023). Comparative analysis of multifaceted neural effects associated with varying endogenous cognitive load. *Commun Biol*, *6*(1), 795. https://doi.org/10.1038/s42003-023-05168-4

Pennycook, G., Tranel, D., Warner, K., & Asp, E. (2019). Beyond Reasonable Doubt. *Neurology and Religion. Cambridge: Cambridge University Press*, 115-129.

Persellin, J. S., Schmidt, J. J., Vandervelde, S. D., & Wilkins, M. S. (2019). Auditor perceptions of audit workloads, audit quality, and job satisfaction. *Accounting horizons*, *33*(4), 95-117. https://doi.org/10.2139/ssrn.2534492

Phillips, F. (1999). Auditor attention to and judgments of aggressive financial reporting. *Journal of accounting research*, *37*(1), 167-189. https://doi.org/doi.org/10.2307/2491402

Raichle, M. E., & Gusnard, D. A. (2002). Appraising the brain's energy budget. *Proc Natl Acad Sci U S A*, *99*(16), 10237-10239. https://doi.org/10.1073/pnas.172399499

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological science*, *9*(5), 347-356. https://doi.org/10.1111/1467-9280.000

Reijula, S., & Hertwig, R. (2022). Self-nudging and the citizen choice architect. *Behavioural Public Policy*, *6*(1), 119-149. https://doi.org/10.1017/bpp.2020.5

Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: how opportunity costs affect strategy selection. *Acta Psychol (Amst)*, *127*(2), 258-276. https://doi.org/10.1016/j.actpsy.2007.05.004

Rose, J. M. (2007). Attention to evidence of aggressive financial reporting and intentional misstatement judgments: Effects of experience and trust. *Behavioral Research in Accounting*, *19*(1), 215-229. https://doi.org/10.2308/bria.2007.19.1.215

Rutkowski, A.-F., & Saunders, C. (2018). *Emotional and cognitive overload: the dark side of information technology*. Routledge. https://doi.org/10.4324/9781315167275

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of aging*, *30*(4), 507-514. https://doi.org/10.1016/j.neurobiolaging.2008.09.023

Salthouse, T. A., Pink, J. E., & Tucker-Drob, E. M. (2008). Contextual analysis of fluid intelligence. *Intelligence*, *36*(5), 464-486. https://doi.org/10.1016/j.intell.2007.10.003

Schouten, J., & Bekker, J. (1967). Reaction time and accuracy. *Acta Psychologica*, *27*, 143-153. https://doi.org/10.1016/0001-6918(67)90054-6

Simon, J. R., & Wolf, J. D. (1963). Choice reaction time as a function of angular stimulus-response correspondence and age. *Ergonomics*, *6*(1), 99-105. https://doi.org/10.1080/00140136308930679

Sisakhti, M., Sachdev, P. S., & Batouli, S. A. H. (2021). The effect of cognitive load on the retrieval of long-term memory: An fMRI study. *Frontiers in human neuroscience*, *15*, 700146. https://doi.org/10.3389/fnhum.2021.700146

Sliwinski, M. J., Mogle, J. A., Hyun, J., Munoz, E., Smyth, J. M., & Lipton, R. B. (2018). Reliability and validity of ambulatory cognitive assessments. *Assessment*, *25*(1), 14-30. https://doi.org/10.1177/1073191116643164

Sridharan, D., Levitin, D. J., & Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc Natl Acad Sci U S A*, *105*(34), 12569-12574. https://doi.org/10.1073/pnas.0800005105

Sterling, P., & Laughlin, S. (2015). *Principles of neural design*. MIT press.

Svanström, T. (2016). Time pressure, training activities and dysfunctional auditor behaviour: evidence from small audit firms. *International Journal of Auditing*, *20*(1), 42-51. https://doi.org/10.1111/ijau.12054

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12*(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4

Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving Decisions About Health. *Wealth, and Happiness*, *3*.

Van Gestel, L., Adriaanse, M., & De Ridder, D. (2021). Do nudges make use of automatic processing? Unraveling the effects of a default nudge under type 1 and type 2 processing. *Comprehensive Results in Social Psychology*, *5*(1-3), 4-24. https://doi.org/10.1080/23743603.2020.1808456

Vannucci, M., Pelagatti, C., Hanczakowski, M., Mazzoni, G., & Paccani, C. R. (2015). Why are we not flooded by involuntary autobiographical memories? Few cues are more effective than many. *Psychol Res*, *79*(6), 1077-1085. https://doi.org/10.1007/s00426-014-0632-y

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67-85. https://doi.org/10.1016/0001-6918(77)90012-9

Wiehler, A., Branzoli, F., Adanyeguh, I., Mochel, F., & Pessiglione, M. (2022). A neuro-metabolic account of why daylong cognitive work alters the control of economic decisions. *Curr Biol*, *32*(16), 3564-3575 e3565. https://doi.org/10.1016/j.cub.2022.07.010

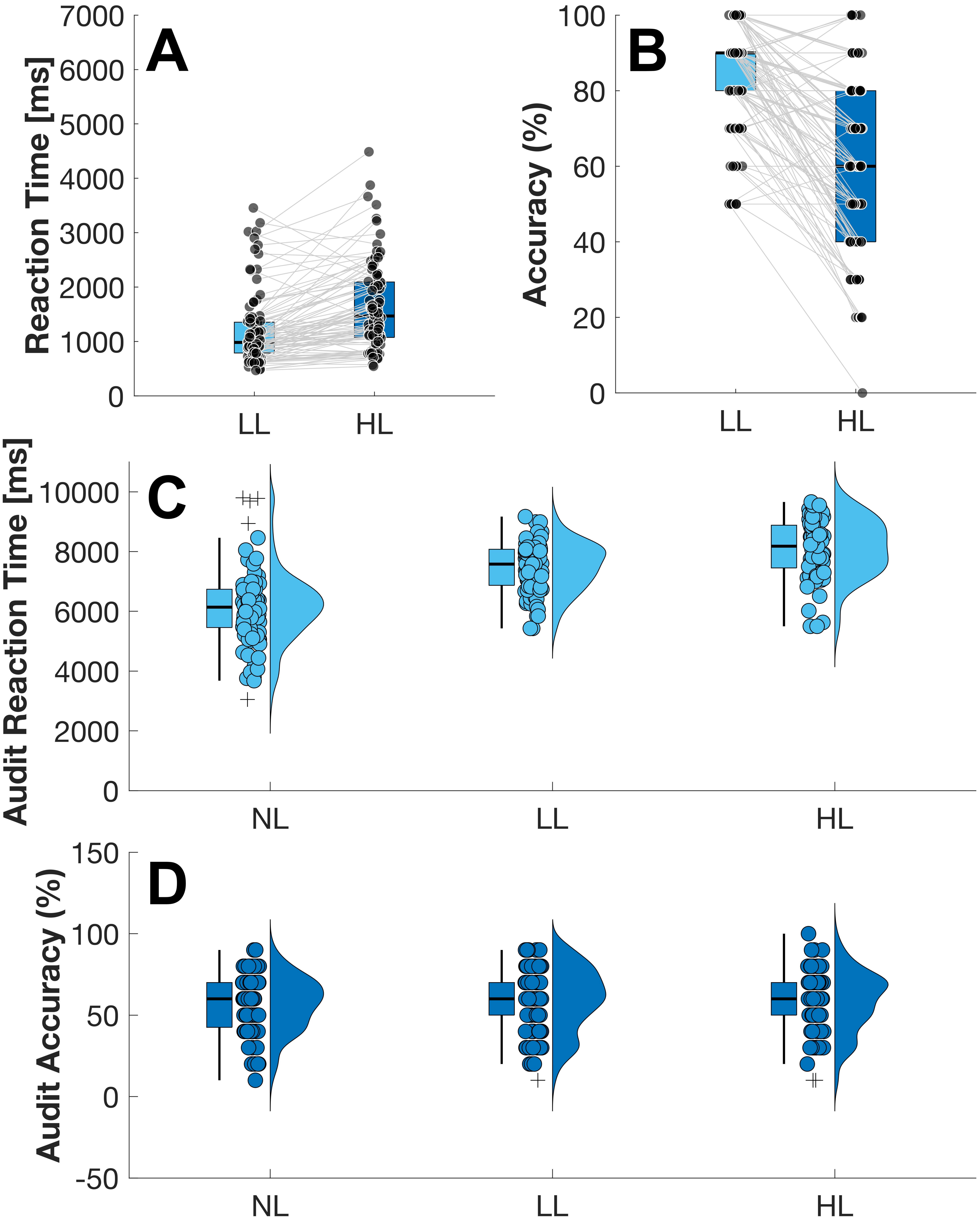
Woodworth, R. S. (1899). Accuracy of voluntary movement. *The Psychological Review: Monograph Supplements*, *3*(3), i. https://doi.org/10.1037/h0092992

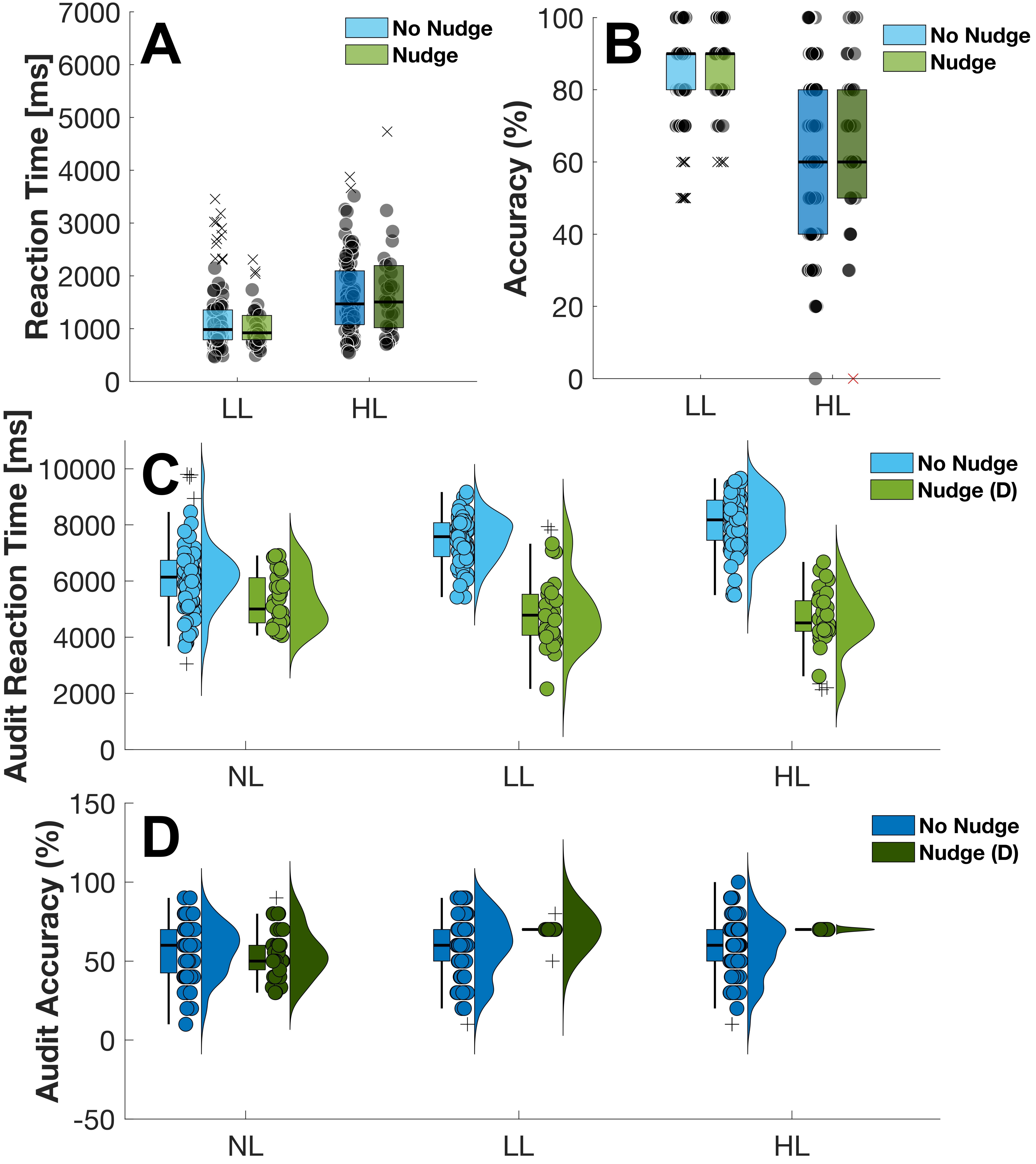
Zhang, B., Zhang, J. X., Huang, S., Kong, L., & Wang, S. (2011). Effects of load on the guidance of visual attention from working memory. *Vision Res*, *51*(23-24), 2356-2361. https://doi.org/10.1016/j.visres.2011.09.008

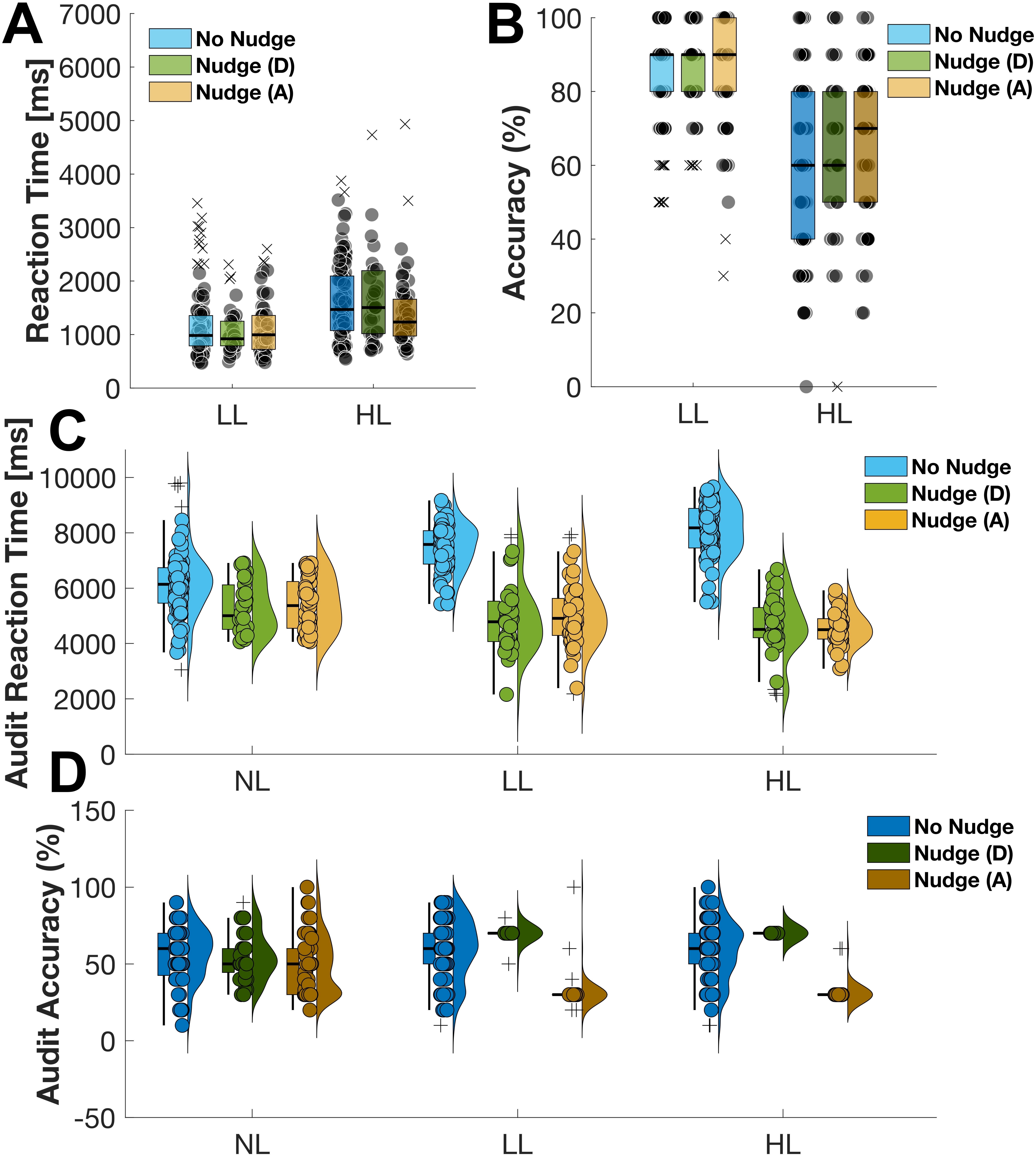
Zhao, X., Chen, A., & West, R. (2010). The influence of working memory load on the Simon effect. *Psychon Bull Rev*, *17*(5), 687-692. https://doi.org/10.3758/PBR.17.5.687

Panel 1 - Exp 1
A
Reaction Time (s)
1.05
1
0.95
0.9
0.85
0.8
0.75
0.7
Minimal Load
Low Load
Medium Load
High Load
B
Adjusted Accuracy (%)
100
90
80
70
60
C
Accuracy (%)
Overall Accuracy
Accuracy + Math
Panel 2 - Exp 2 to 3b
No Nudge
Nudge (D)
Nudge (A)
A
B
C
Reaction Time (s)
10
8
6
4
2
0
NL
LL
HL
D
E
F
Accuracy Audit (%)
150
100
50
0
-50
Panel 3 - Exp 2 to 3b
No Nudge
Nudge (A)
Nudge (D)
A
Reaction Time (s)
B
Accuracy Audit (%)
NL
LL
HL
Panel 4
Exp 2
Exp 3a
Exp 3b
A
B
C
Reaction Time (s)
D
E
F
Adjusted Accuracy (%)
LL
HL
No Nudge
Nudge (D)
Nudge (A)

A
Objective Scepticism
0
1
2
3
No Nudge
Nudge (A)
Nudge (D)
NL
LL
HL
B
True Positives (%)
C
False Negatives (%)
D
False Positives (%)
E
True Negatives (%)
0
10
20
30
40
50
60
70
80
90
100
NL
LL
HL
No Nudge
Nudge (A)
Nudge (D)


A
Reaction Time [ms]
7000
6000
5000
4000
3000
2000
1000
0
LL
HL
B
Accuracy (%)
100
80
60
40
20
0
LL
HL
C
Audit Reaction Time [ms]
10000
8000
6000
4000
2000
0
NL
LL
HL
D
Audit Accuracy (%)
150
100
50
0
-50
NL
LL
HL


A
Reaction Time [ms]
7000
6000
5000
4000
3000
2000
1000
0
LL
HL
No Nudge
Nudge
B
Accuracy (%)
100
80
60
40
20
0
LL
HL
No Nudge
Nudge
C
Audit Reaction Time [ms]
10000
8000
6000
4000
2000
0
NL
LL
HL
No Nudge
Nudge (D)
D
Audit Accuracy (%)
150
100
50
0
-50
NL
LL
HL
No Nudge
Nudge (D)


A
Reaction Time [ms]
7000
6000
5000
4000
3000
2000
1000
0
LL
HL
No Nudge
Nudge (D)
Nudge (A)
B
Accuracy (%)
100
80
60
40
20
0
LL
HL
No Nudge
Nudge (D)
Nudge (A)
C
Audit Reaction Time [ms]
10000
8000
6000
4000
2000
0
NL
LL
HL
No Nudge
Nudge (D)
Nudge (A)
D
Audit Accuracy (%)
150
100
50
0
-50
NL
LL
HL
No Nudge
Nudge (D)
Nudge (A)

Retention
1s
Encoding
1s
1s
5s
Recall
Self-paced
Exp. 1
Basic Mathematical Operations
Self-paced
Load Levels
Low Load
Medium Load
High Load
Exp. 2
Audit Task No Nudge
Exp. 3a
Audit Task Nudge on "Disagree"
Exp. 3b
Audit Task Nudge on "Agree"