# Metacognitive Switch Cost:

# Metacognitive Perspective on Task Switching Effect

Yaara Gur Esh

**Research Thesis**

In partial fulfillment of the requirements for the degree of Master of Science in Behavioral and Management Science (Cognitive Psychology and Human Factors Engineering).

The research thesis was done in the faculty of Data & Decision Making,

under the supervision of Professor Rakefet Ackerman.

Submitted to the Senate of the Technion - Israel Institute of Technology

Shvat 5785, Haifa, February 2025

**Table of Content**

**List of Figures**

**List of Tables**

**Abstract**

In dynamic task environments, problem-solving demands continuous self-assessments of confidence in potential solutions, which play a crucial role in guiding effective regulation of thinking efforts. In this research we examined the interaction between metacognitive monitoring of confidence and switching between tasks - an unexplored area. For this purpose, we used the Bird's Eye View of Cue Integration (BEVoCI) methodology (Ackerman, 2023). This method allows examining at once the extent to which multiple heuristic cues predict success and confidence. Across two experiments, participants solved two problem types, pairing a primary non-verbal (geometric) reasoning task, for which we examined the switching effect, with a secondary task that induced the switching effects on the primary one. Experiment 1 featured three task switching conditions: separated-tasks, predictable switching, and unpredictable switching, alternating between the non-verbal task and a verbal task. Experiment 2 included the separated-tasks condition and the unpredictable switching condition, focusing solely on non-verbal tasks and introducing task color as a moderating factor. The results revealed interactions between task-switching (reflecting varied cognitive load), task-design (i.e., stimuli color), and heuristic cues in predicting monitoring processes. These findings are interpreted through the interplay between cognitive load and heuristic cues, which may inform the selection of strategies that enhance self-regulation, predict performance, and shape metacognitive performance during complex reasoning tasks.

# 1. Introduction

In the ever-changing landscape of life, task switching is a constant presence. We are tasked with a multitude of cognitive challenges, including processing information, mastering time-management, inhibiting distractions, processing data, and making decisions. When faced with multiple challenging tasks, human behavior exhibits a high degree of adaptability and flexibility in response to the ever-changing environmental demands (Kiesel et al., 2010).

Metacognition plays a significant role in adapting and directing our cognitive processes (Nelson & Narens, 1990), for example, decision-making. In essence, metacognition is the "top manager" of cognitive functioning and responsible for regulating a range of functions, such as setting goals, allocating time for problem solving, strategy utilization, and next step decisions. In practice, metacognition is not separated from cognition, but rather an integral part of higher-order cognitive processes, serving as the quality-control mechanism of one's mental functions (Fiedler et al., 2019).

So far, the vast majority of metacognitive research has focused on a single task type, typically involving multiple items of the same type, such as a set of similar problem-solving items (e.g., Syllogisms: Bajšanski et al., 2014; Raven matrices: Raven, 2003). The current study aims to investigate the impact of task switching on metacognitive processes. The following sections introduce task-switching frameworks, metacognitive monitoring and measures, provide an overview of the current study and its hypotheses, finally, describe the planned experiments.

## 1.1. Task-switching

Task switching refers to the process of task-set reconfiguration, adapting to changing goals, and reassigning cognitive resources from one task to another (Hazeltine, 2024; Monsell, 2003). Task shifting often involves shifts of attention and updating the contents of working memory (e.g., recalling a number we memorized a few seconds ago). These shifts may slow down and harm performance (Hazeltine, 2024). For studying these costs, task-switching paradigms have been developed (Vandierendonck et.al., 2010). The following sections review these two aspects of task switching research.

To cope with a single challenging task, people constantly monitor their progress and act to achieve their goal. However, when a task switch is required, performance usually drops relative to performing each task in isolation. This is known as ‘switch cost’ and it frequently harms efficiency (Gonzalez-Gomez et al., 2023), reflected by slower task performance, increased error rates, and decreased focus on each task’s relevant information (Kiesel et al., 2010; Vandierendonck et al., 2010).

Switch-cost often reflects the operation of processes involved in cognitive control (Monsell, 2003). Task set reconfiguration - reflected in the cognitive control processes - is often cited as the de facto explanation for switch costs, reflecting the time needed to update and reallocate a task-set in working memory to ensure stimuli elicit the required responses for the currently handled task. Psychologist Arthur T. Jersild first described the switch-cost effect in his 1927 work, Mental Set and Shift. As a result of his novel task-switching paradigm, Jersild discovered that the amount of time it takes for the brain to react is slower when it alternates between different tasks than when it repeats the same tasks.

The primary objective of the current study is to investigate the metacognitive implications associated with switching between tasks. As far as we know, the metacognitive implications of the task-switching paradigm remain unexplored. In terms of metacognitive processes, although prior studies focused on multitasking or dual-tasks rather than task switching, they underscore the importance of investigating cognitive and metacognitive performance under cognitive load, such as task switching. For example, prior research has demonstrated that individuals can express the expected performance costs in their confidence ratings when faced with multitasking (Finley et al., 2014). Additionally, the effect of cognitive overload on production deficiency in metacognitive activities - that was investigated using a dual-task paradigm - has been shown to result in cognitive overload, which impairs metacognitive processes (Sannomiya & Ohtani, 2015). A study investigating the dual processes of reasoning found that reasoning performance decreased under cognitive load (De Neys et al., 2006). Gopher et al. (2022), using the Breakfast Task (developed by Craik & Bialystok, 2006), showed that high demands on executive control related to task difficulty (e.g., higher-order thinking skills such as organizing and planning) impact priority setting, leading to performance impairments such as longer response times and increased error rates. Notably, multitasking, or “dual-tasking”, is described in the scientific literature as performing two tasks simultaneously - for example, driving and talking to a passenger, or walking while interacting with the mobile (Hazeltine et al., 2006). In contrast, task-switching involves performing more than one task in short temporal succession, while switching back and forth between them, that is, performing only one of them in each specific moment.

Task switching research is typically done by two main paradigms: predictable and unpredictable sequences. The predictable task sequence is known as the "Alternating runs" paradigm (Rogers & Monsell, 1995). According to this paradigm, the sequence patterns (e.g., length and switch points) are known to the participants in advance, and the task switching occurs systematically, and after a consistent number of trials involving the same task (e.g., switch every second trial; AA, BB, AA, BB). As an alternative to the predictable sequences, a "Task-cuing" paradigm with unpredictable sequences has been developed (e.g., Meiran, 1996). In this paradigm, the order of the task sequence and the switch points are random and thus unexpected. Similar to Rogers and Monsell's study (1995), the unpredictable sequences revealed a robust switch-cost in switch-trials compared to separated tasks (Gopher, 2000).

Although prior studies revealed that compared to unpredictable trials, predictability enhances preparation for the upcoming task (Nicholson et al., 2006) and improves processing fluency - the ability to navigate and optimize cognitive process (Serra & Shanks, 2023), it is important to note that the switch-cost still remains substantial in the predictable trials, even when the task is quite simple and the task's switching is entirely predictable (Koch, 2003). In their work, Rogers and Monsell (1995) compared performance of predictable trials to separated-tasks (termed "Repetition" task-switch trials., AAAA, BBBB; Task B performed after Task A or vice versa). The major finding was impaired performance in the predictable trials compared to separated-tasks.

The present study compared the three switching sequences: separates-tasks, predictable and unpredictable conditions in their effect on metacognitive monitoring. To the best of our knowledge, no prior research investigated metacognitive monitoring concerning switching between tasks, a gap we aim to fill in by this study.

### 1.2. Metacognitive processes

Nelson and Narens' (1990) framework of metacognition introduces the metacognitive processes of monitoring, which involves assessing one's own cognitive performance. By this framework, people set a target that reflects their current motivation to succeed and during task performance - monitor their progress relative to this target level. Traditionally, studies investigating metacognitive monitoring processes evaluate participants' performance using meta-reasoning tasks. The meta-reasoning framework focuses on problem solving, in which participants assess (monitor) their performance regarding the chance of their solutions to be correct (Ackerman & Thompson, 2017). A monitoring type that is used in metacognitive research is the judgement of confidence, which was employed in the current study. Confidence reflects the chance of a considered answer being correct during a contemplation process and after a response is provided (e.g., Ackerman, 2023).

An important aspect in metacognitive research is discovering heuristic cues underlying metacognitive monitoring. Koriat's (1997) cue-utilization approach is well-established in the metacognitive research, suggesting that various types of cues underly monitoring the chance to succeed while performing a task. Given that monitoring is based on cues, rather than on actual indicators of success, these cues potentially mislead metacognitive judgments. Metacognitive research has identified numerous cues that may bias confidence judgments (e.g., familiarity with question terms, see Ackerman, 2019, for a review). In her work on the Bird's Eye View over Cue Integration (BEVoCI), Ackerman (2023) demonstrated that different cues have varying degrees of predictive values for success in the task, and that contextual factors (e.g., task design) may affect how cues are utilized. The current study investigated the extent to which cue validity (predicting success) and cue utilization (predicting confidence) are affected by task-switching conditions.

## 2. Overview of the experiments

The current research consists of two experiments; each one comprised different task-switching conditions. Within each condition, participants performed two distinct reasoning tasks, which were presented as equally important to the participants, while in terms of research one was the primary task, which generated the main data for analysis, and a secondary task, used to create the task-switching effects on the primary task. Both experiments involved similar design and instructions. The primary task remained the same in both experiments, while the secondary task differed between the experiments. In all conditions, the participants knew in advance about the two tasks they are going to perform and went through practice items before starting the solving phase. The participants also knew in advance the total number of items they are going to solve of each task type.

### 2.1. Task-switching manipulation

Both experiments employed a set of task-switching conditions, with Experiment 1 featuring three switching conditions while Experiment 2 included two switching conditions. The conditions included in Experiment 1 were: (1) Separated-tasks: task switching occurred once after the completion of one whole task. (2) Predictable switching: switching every constant number of items. That is, the switch points and so are the sequences length were known to the participants in advance (e.g., Rogers & Monsell, 1995). (3) Unpredictable switching: switching after an unknown number of items (e.g., Meiran, 1996). In Experiment 2, only the separated-tasks and unpredictable switching conditions were applied. In both experiments, each condition incorporated two different task orders, where participants started with either one of the two tasks. These orders were counterbalanced between participants.

### 2.2. Experimental design

To measure the switch-costs, reasoning tasks that require immediate responses should be employed, rather than memory tasks where recall typically occurs in later trials. This approach is necessary because response selection, as noted by Yamaguchi (2024), is a critical factor in the generation of switch costs.

The present study incorporated three reasoning tasks. Both experiments included the same primary task, the *Comparison of Perimeters* task (CoP; Ackerman, 2023), involving non-verbal multiple-choice riddles based on geometric shapes (see Figure 1). This task provided an opportunity to examine metacognitive aspects of visual cognition, a key component of daily tasks such as engineering and design, which are rarely examined from this perspective. Additionally, each experiment included a secondary task alongside the primary CoP task.

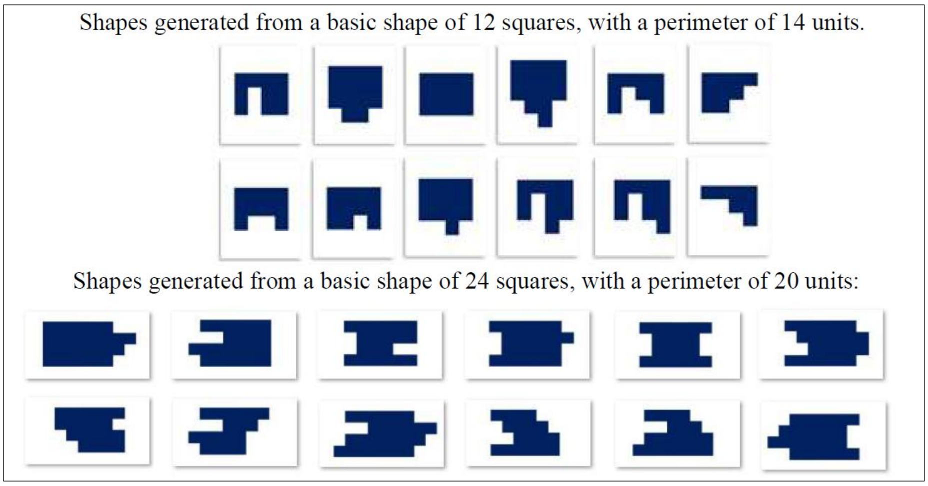
Experiment 1 paired with the non-verbal CoP task a verbal task, the *Compound Remote Associate* (CRA, Bowden & Jung-Beeman, 2003), with an open-entry answer, adapted from Ackerman et al. (2020). These two clearly distinct reasoning tasks feature different goals and cues. Experiment 2 paired with the CoP another non-verbal multiple-choice task, the *Missing Tan Task* (MTT), using geometric silhouettes, adapted from Ackerman (2023) and based on the Tangram game (https://en.wikipedia.org/wiki/Tangram). Although the tasks in Experiment 2 were different, they shared the focus on visual processing of geometric shapes. All tasks

(CoP, CRA, MTT) included a judgment of confidence elicited immediately after providing each problem solution. Examples for both secondary tasks (i.e., CRA, MTT) are presented in the Material sections of each experiment.

We anticipated that the noticeable differences between the tasks in Experiment 1 would serve as an extrinsic cue (Koriat, 1997), clearly signaling when task switching occurs. Prior research has shown that when stimuli differ substantially between tasks (e.g., words vs. shapes), the mental switch is less demanding (Hazeltine, 2024; Waszak et al., 2003). In contrast, switching between tasks with common elements (e.g., similar stimuli, such as shapes of the same color) - as designed in Experiment 2 – may significantly increase switch costs (e.g., longer response times), as it requires additional attention to recognize the switch (Waszak et al., 2003). Therefore, Experiment 1 focused on switching between substantially different tasks, both visually and cognitively (i.e., different visual stimuli and goals), ensuring that task switches were clearly signaled. This design aimed to isolate the effects of switching itself while minimizing confounding influence from stimulus overlap (e.g., similarities between task stimuli). The effects of stimulus overlap are at the focus of Experiment 2.

**Figure 1**

*Stimuli pool for the Comparison of Perimeters (CoP) task (Ackerman, 2023).*



The global dependent variables in both experiments were the classic metacognitive research measures (see Ackerman et al., 2016, for details regarding these measures) – namely success (right or wrong) and confidence. Ackerman (2023) found that instructional design affects the weight balance among cues that predict success and underlie confidence. We adapted the cues used by her for the CoP task to examine how task switch affects cue validity and cue utilization.

The predictive independent variables were the levels of the various cues examined for each task (e.g., edge differences ranged between 0 and 6). Each cue produced a beta coefficient when predicting success and confidence, by the hierarchical multiple regression models, using the BEVoCI method. We distinguish between task-based cues derived from each item and common across participants, while response time reflects the variability across participants in dealing with each item. The following heuristic cues were hypothesized to predict success and confidence:

A. *Response Time*. The time elapses from when a problem is displayed until participants select their response. Response time is the most commonly examined cue in metacognitive research, and often used as a proxy for fluency, reflecting the individual effort invested in each item (Thompson et al., 2013). The theorizing behind this is that response time reflects the participant's individual experience of the item as easy or difficult, which informs how they infer their chance for success (Unkelbach, 2007). Including response time together with task-based cues in regression analyses allows isolating the unique contribution of the individual-item interaction above and beyond task characteristics that are common across participants. That is, response time is the only cue that reflects the unique variance for each participant when interacting with each item.

B. *Task-based cue 1: Perimeter–area congruency*. While asking participants to compare the perimeters of two shapes, the shape area has been found to be highly misleading (Ackerman, 2023; Stavy & Babai, 2008). Congruency exists when the shape with a longer perimeter in the pair also has a larger area. Incongruency between perimeters and areas exists when one shape has a longer perimeter than the second shape, with equal or smaller area, and vice versa. The question regarding cue utilization is to what extent confidence ratings reflect the difference in task difficulty that stems from misleading cues like perimeter-area congruency. Ackerman (2023) found that confidence was entirely insensitive to this cue, despite a success rate difference of approximately 41% between congruent and incongruent shape pairs.

C. *Task-based cue 2: Basic shape area*. The basic rectangular shape is composed of 12 identical small squares arranged in a 3 X 4 pattern, with squares removed or added on one side only. Following Ackerman (2023) who aimed to increase task complexity, half the items consisted of 24 squares in a 6 X 4 pattern arrangement, and squares were removed or added on two sides (see Figure 1). Ackerman (2023) found confidence overly sensitive to cue variations despite success rate differences being significant but smaller.

D. *Task-based cue 3: Difference in edges*. An edge is a straight side, forming part of the perimeter between two corners of a geometric shape. Shapes with a greater number of edges (or corners) often appear more complex. However, a higher number of edges does not necessarily correspond to a longer perimeter. For example, as shown in Figure 1, the rightmost shape on the top row has 8-edges, compared to a basic rectangle with 4-edges of the same length and width. Despite having more edges, the perimeter of the 8-edges shape is identical to that of a basic 4-edges rectangle. Shape pairs with a smaller difference in the number of edges are more difficult to distinguish by perimeter length compared to pairs with a larger difference in edges. In Ackerman (2023), confidence ratings showed only slight confidence sensitivity to the substantial effect of differences in edges on success.

In addition to the examination of potential differential effects of task switching conditions on the predictive power of heuristic cues, for completeness and comparison with prior research, two additional measures were examined in each experiment: Calibration and Resolution (Ackerman et al., 2016):

I. *Calibration (overconfidence).* The notion of metacognitive calibration entails the correspondence between the absolute numerical value of metacognitive judgments and actual success rates (Keren, 1991). Calibration is determined by the delta between the mean judgement and the actual success rates. A well-calibrated participant will show a small delta (e.g., $\Delta = 0.5$). Conversely, when the delta is substantial, implying a disparity between the mean judgment and actual performance (e.g., $\Delta = 25$), a positive delta reflects overconfidence, in which participants evaluate their success above their actual success rates, while a negative delta (when success rate surpasses mean judgments) indicates under-confidence (Fleming & Lau, 2014; Tobler & Kapur, 2023). Well-calibrated monitoring is important because it is by this subjective assessment that people allocate their thinking efforts (e.g., Ackerman & Goldsmith, 2011).

II. *Resolution.* Resolution refers to distinguishing between correct and incorrect answers (Fleming & Lau, 2014). In classic metacognitive research, a common index of resolution is the Goodman-Kruskal Gamma correlation between the metacognitive judgment and the correctness of each individual item, calculated *within* individuals (Nelson, 1984). A strong positive Gamma correlation indicates a high level of resolution

(Vuorre & Metcalfe, 2022) that is a prerequisite for effective control decision (e.g., restudy choices, Thiede et al., 2003).

In sum, this study seeks to assess the extent to which people monitor their performance accurately, despite the cognitive load associated with task switching. Specifically, our focus was on the differences in cue validity and in cue utilization between the three task-switching conditions, namely, separated-tasks, predictable and unpredictable switching. By this analysis we aimed to determine which switching condition allows participants to perform more effectively (i.e., minimal switch cost) and shed light on the underlying metacognitive processes contributing to this outcome.

Prior research suggests that switching costs are larger in unpredictable task sequences compared to predictable ones (e.g., Schneider & Logan, 2006a). Thus, we expected worse success, lower confidence, and longest response times in the unpredictable switching condition (i.e., where the cognitive load is increased), with the separated-tasks condition demonstrating the least switch cost. In terms of cue validity and utilization, given this study is the first to consider it, we explored the effects of the switching condition on the extent to which task-based cues and response time predict success and confidence.

Differences in success and confidence were expected to be associated with variations in task-based cues and response time. Ackerman (2023) has shown that confidence and success are both sensitive to these heuristic cues. Specifically, we expected to see effects of the switching conditions on the predictive strength (i.e., beta coefficient; see Results and Discussion in Section 3.2) of the considered task-based cues (e.g., incongruent vs. congruent shapes in the Perimeter-area congruency cue) and in response time. We anticipated that in both predictable and unpredictable switching groups, the gap would stem from lower success and confidence (or longer response times) in the more challenging shape pairs , whereas in the separated-tasks group, the gap would result from relatively higher success and confidence (or shorter response times) in the easier shape pairs in the stimuli set. To measure the differences in the relationship between success and confidence by delving into the effects of the task-based heuristic cues and response time, we used the BEVoCI method proposed by Ackerman (2023). A key advantage of the BEVoCI is its ability to simultaneously examine multiple cues while controlling for all other variables.

## 3. Experiment 1

The aim of this experiment was to document the metacognitive switching cost when the tasks were inherently and visually substantially different from each other. It included the non-verbal CoP task and the verbal CRA task with the three switching protocols described above, separated-tasks, predictable switching, and unpredictable switching.

Within the separated-tasks group, half of the participants were included to measure the extent to which the current sample replicates Ackerman's (2023, Experiment 1a) findings within the primary task (CoP) results. The replication group termed as the Baseline condition (see Table 2). The other half of the participants, with the CoP appearing entirely after all the CRA items, and thus performed the CoP after one task switch.

### *3.1. Method*

#### *3.1.2. Participants*

A power analysis using G*Power (Faul et al., 2007) revealed that for a hierarchical multiple regression model with three groups, a sample size of 128 participants (approximately 40 per group) would be required to achieve a power of 0.8 with an effect size of 0.3, considered a medium effect size based on established guidelines (Cohen, 1988). Similar studies that revolved around metacognitive measures with a similar sample size (however, with no switching designs) showed similar effect sizes with similar or somewhat larger sample sizes (e.g., judgement-time correlation, around $N = 50$ per group, CRA reasoning task: Ackerman, 2014; Confidence judgement, sample size between 50-80 per group, CoP reasoning task: Ackerman, 2023).

Participants' inclusion criteria were: (i) Participants had to successfully complete at least three attention-verification items out of four (see Materials). (ii) A dummy statement included in a self-report after the main tasks (see Materials) served as another attention check. (iii) Focus on the task window had to be higher than 75% of the time spent on the task. (iv) At the item level, response time for the CoP had to be less than thirty seconds (otherwise it was assumed the participant was distracted), and focus time on the window had to be greater than 50% of the time spent on the item. (v) For those who responded particularly quickly (less than 2SD of the sample), the success rate of the CoP had to be higher than chance (33%). (vi) Confidence had to show some variability unless success was higher than 90% and confidence was 100% for all items. (vii) Participants had to provide at least 85% usable items after the item screening. Participants were excluded only if they violated two or more of these criteria.

We recruited 210 native English-speaking adults from Prolific Academic. Following the pre-defined screening criteria, eighteen participants were excluded: seven due to low number of valid items and/or low focus (criteria iii, vii), seven due to low confidence variability (criterion vi), and four due to attention verification failures (criteria i, ii), leaving 192 participants for the data analyses ($M_{age}$ = 37.93, *SD* = 13.31, 52% females).

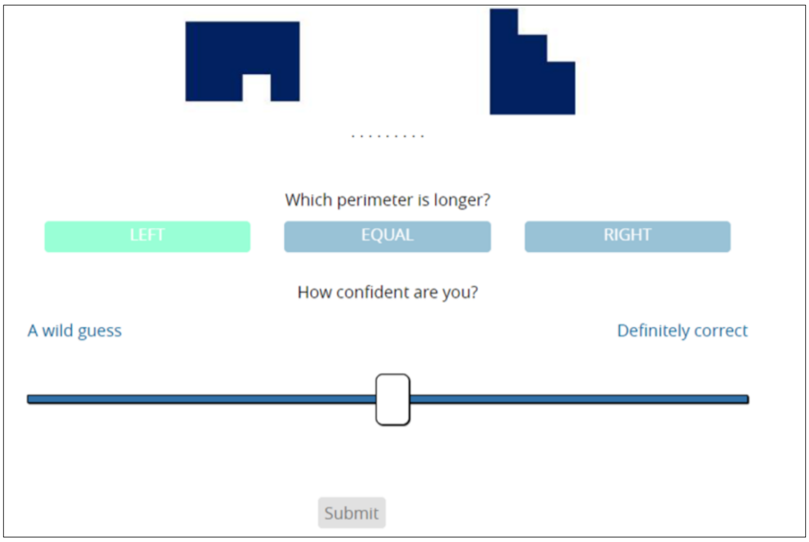
Applying the task-switching manipulation, participants were randomly assigned to one of the conditions: 64 to the separated-tasks, 65 to the predictable switching, and 63 to the unpredictable switching. Participants were asked to devote about 30 minutes without interruptions. The compensation for participating was 2.5 GBP.

### *3.1.3. Materials*

*Comparison of Perimeters (CoP).* This task included 21 geometric problems. Each problem included two shapes, generated from a basic shape derived from a basic rectangular of 12 squares (3 X 4) with perimeter of 14 units, or a basic shape of 24 squares (4 X 6) with a perimeter of 20 units (see Figure 1). The participants' task was to determine which shape, if either, had a longer perimeter (Figure 2). The included problems were a subset of the stimuli used by Ackerman (2023), representing high difficulty variability (29% - 100% success rates).

**Figure 2**

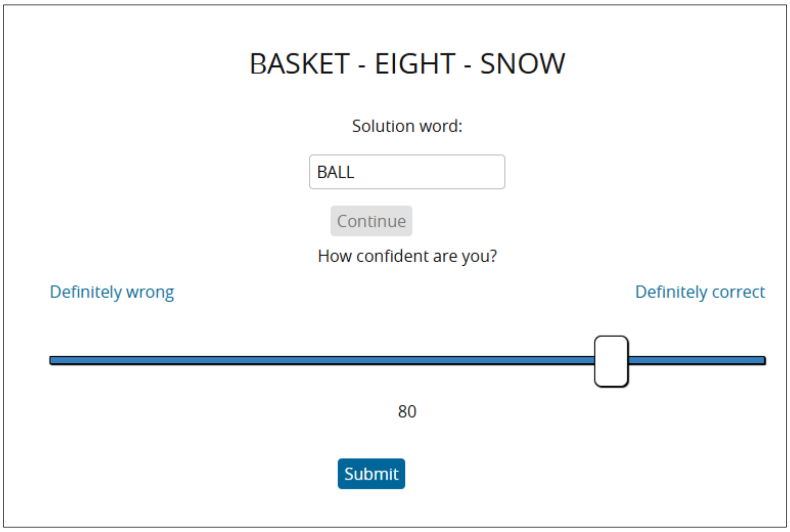
*Primary task: Comparison of Perimeters (CoP) task.*



*Compound Remote Associate (CRA).* The CRA task included sets of three words. Participants were required to find a fourth word that, when combined with each of the three, formed a compound word or common two-word phrase (see Figure 3). For example, the words BASKET, EIGHT, and SNOW have the solution word BALL, forming BASKETBALL, EIGHTBALL, and SNOWBALL. The task included 21 problems, with success rates ranging from 20% to 80% (Lauterman & Ackerman, 2023).

**Figure 3**

*Experiment 1's secondary task: Compound Remote Associates (CRA) task.*




Following the 42 items of both main tasks, participants responded to several self-report questions on a 1‑7 scale. The last among them was a dummy statement used as an attention verification criterion (v above); the question instructed participants to ignore the statement and respond with a specific number on the scale.

*3.1.4. Procedure*

Participants took part in the experiment online. Within each condition group (i.e., separated-tasks, predictable, or unpredictable switching), participants were randomly assigned to one of two task-orders, with counterbalancing either starting with CoP or CRA.

First, the tasks instructions were presented, along with examples and practice. The instructions also pointed that the tasks would be switching automatically, while the groups of the separated-tasks and predictable switching also knew the sequences' length. In all three conditions, participants rated confidence immediately after clicking an answer. Judgements were elicited by a 33-100% scale for CoP riddles (33% was the chance level of a wild guess) and 0-100% scale for CRA riddles. Experiments had no time limitation and response time was recorded for each item.

### *3.2. Results and discussion*

Descriptive results of the CoP task, which was the primary task, are presented in Table 1. Probing into the overall results, a one-way Analysis of Variance (ANOVA) comparing the three switching conditions revealed no significant differences in terms of both cognitive and metacognitive performance. This finding is at odds with prior literature, which suggested that unpredictable switching was expected to harm performance.

**Table 1**

*Experiment 1: Means (SD) of cognitive and metacognitive measures for the Comparison of Perimeter (CoP) task.*

| Condition | Baseline Condition | Separated-Tasks | Predictable Switching | Unpredictable Switching |
|---|---|---|---|---|
| Success rates (%) | 64.9 (14.2) | 62.5 (13.3) | 68.9 (13.7) | 66.2 (14.3) |
| Confidence (%) | 76.7 (8.67) | 79.7 (9.28) | 77.5 (8.94) | 77.0 (8.67) |
| Overconfidence | 11.8 (18.3) | 17.2 (14.4) | 8.56 (13.1) | 10.8 (16.0) |
| Response Time (sec.) | 12.8 (11.8) | 12.6 (7.71) | 16.2 (11.2) | 15.4 (12.7) |
| Resolution (gamma) | .26 (.33) | .29 (.40) | .33 (.34) | .29 (.33) |

For examining the main hypotheses, we applied the BEVoCI methodology (Ackerman, 2023) that allows looking into the relative contribution of each heuristic cue in comparison across the task switching conditions. Before conducting the analysis, all dependent variables were standardized within participant. We then used hierarchical multiple regression models to predict both success and confidence by the heuristic cues listed above. The models were implemented in R 4.1.0, using the *nlme* package (Pinheiro et al., 2019). Within each group, Level 1 represented the items, while Level 2 represented the participants. The beta coefficients and their significance are presented in Table 2. Notably, task-order did not have a significant predictive value for neither success or confidence ($p > .20$). When comparing models including task-order and those from which it was excluded, its inclusion did not change the significancy or the betta coefficients of other predictors. Thus, we waived it as a potential cue and included it in the model as a fixed effect, controlling for its variance.

Starting with the baseline condition, the results indicate a partial replication of Ackerman's (2023) findings. Most predictors' effects, in terms of predicting both success and confidence, aligned with Ackerman’s (2023) results, including both coefficients and significance. However, one notable observation was the weak and insignificant effect of perimeter-area congruency on success in this study. We refer to this finding as the reference point for the present study, as this group took part in random allocation with the others.

**Table 2**

*Experiment 1: Hierarchical Multiple Regression predicting success and confidence - Beta coefficients within each group.*

| Condition | Ackerman, 2023 Experiment 1a | | Baseline Condition of the current study – Ackerman's 2023 replication | | Current study Switching Conditions: Separated-tasks | | Predictable Switch | | Unpredictable Switch | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Success | Confidence | Success | Confidence | Success | Confidence | Success | Confidence | Success | Confidence |
| Perimeter-area congruency | .41 *** | .01 | .08 ($p$ = .053) | .06 | 0 [a] | .01 | .07 * [ab] | .02 | .12 *** [b] | .03 |
| Basic shape area | -.08 *** | -.19 *** | -.06 | -.20 *** | -.07 | -17 *** | -.05 | -.15 *** | -.07 * | -.12 *** |
| Difference in edges | .20 *** | .12 *** | .21*** | .13 *** | .35 *** [b] | .16*** [b] | .17 *** [a] | .13 *** [ab] | .19 *** [a] | .07 ** [a] |
| Response time | .02 | -.07 *** | -.08 | -.11* | -.04 | -.04 [ab] | .01 | -.01 [a] | 0 | -.14 *** [b] |

*Note*. Significance of a cue as a predictor, $^{*}p < .05$, $^{**}p \leq .01$, $^{***}p \leq .001$

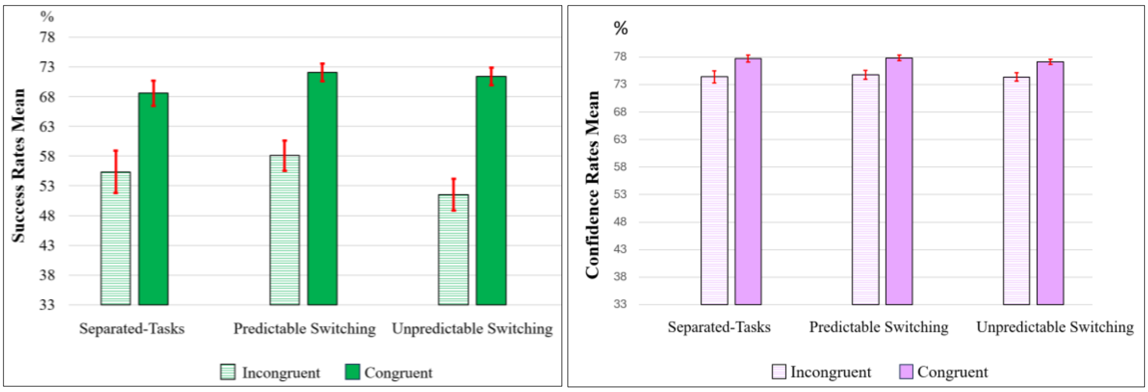
Gray font: Insignificant betta coefficient.

[a,b]: Different letters signify a significant pairwise difference between switching conditions in success or confidence, $p < .05$.

Looking at Table 2, while comparing the three conditions, reveals several observations. Perimeter-area congruency predicted overall, across the conditions, a significant effect on success rates only within the predictable and unpredictable switching groups, but not in the separated-tasks condition. In contrast, confidence was consistently blind to the congruency within all groups. Nevertheless, as can be seen in Figure 4, across all groups, both success rates and confidence were higher for congruent pairs than for incongruent pairs. When examining the task-switching effect, the largest gap in success rates between congruent and incongruent pairs observed in the unpredictable switching group, significantly differing from the separated-tasks group. The difference in beta coefficients for perimeter-area congruency between the groups stems from the reduced success rates on the more challenging shapes (incongruent pairs) in the unpredictable switching group (Figure 4 left panel). This finding highlights the moderating role of task switching on success rates, indicating that success is influenced not only by the predictive heuristic cue but also by the cognitive load imposed by a specific task-switching design. This effect is particularly pronounced for challenging stimuli (i.e., incongruent pairs), while performance remains consistent for easier shapes (i.e., congruent pairs). In contrast, confidence (Figure 4, right panel) remained blind to effects of the switching manipulation and congruency differences on success.

**Figure 4**

*Experiment 1: Success and Confidence by Perimeter-Area Congruency across groups. Left Panel: success results, Right Panel: confidence.*




*Note*: Error bars represent ± 1 SE.

Basic shape area significantly predicted success only in the unpredictable switching condition. In contrast, for confidence, this cue was a consistent predictor across all switching groups, aligning with the baseline condition. Notably, there were no significant group differences (Table 2), indicating that task switching leaves the relationship between basic shape area and both success and confidence unaffected. The combined findings of perimeter-area congruency and basic shape area expose that confidence was not just stable (Figure 4, right panel) but was sensitive to some cues but not for others.

Difference in edges had an overall, across all groups, a significant effect on success and confidence, showing a positive trend as the difference in edges between the two presented shapes was bigger (i.e., reduced cognitive challenge), as illustrated in Figure 5. This pattern consistently replicated the baseline condition. When examining the task-switching effect, the largest gap in success rates between shapes with zero edge differences (most challenging) and those with 4+ edge differences (easier pairs) was observed in the separated-tasks group, compared to the predictable and unpredictable switching groups. Additionally, difference in edges predicted the largest gap in confidence between the most challenging pairs and the easier ones in the separated-tasks group, compared to the unpredictable switching groups. The difference in the success beta coefficients for the edge differences between the groups (see Table 2) stems from the lower success rates observed in the most challenging shapes within the separated-tasks group, which also showed the most substantial improvement in success rates as edge differences increased (i.e., easier shapes; see Figure 5). Although this finding is somewhat in line with our prediction, according to which a performance improvement is more associated with low switch-cost, the unexpectedly low success rates for challenging pairs in the separated-tasks group were surprising still. To explore this further, a one-way ANOVA was conducted across the groups on the success rates of shapes with no edge difference, and of shapes with 4+ edge difference. The ANOVA unequivocally revealed a significant effect for the success of no edge difference shapes, while no significant effect for success of 4+ edge difference shapes between the groups was found (Table 3). A TukeyHSD post-hoc revealed that the differences in the success of 4+ edge difference shapes stemmed from the gap between the separated-tasks and the predictable switching group, while the success rates pattern between the predictable and unpredictable switching groups remained consistent. These findings highlight that although the separated-tasks group demonstrated the most significant improvement in success rates between challenging and easier shape-pairs, they experienced the highest switch-cost when presented with challenging stimuli. As for confidence, the difference in the beta coefficients for difference in edges between the groups (Table 2), stemmed from a modest increase in confidence for easier shapes in the separated-tasks group (Figure 5, solid purple line). In sum, the findings regarding difference in edges suggest that the separated-tasks group, characterized by lower levels of cognitive load, demonstrated a more pronounced increase in both success and confidence when presented with easier items, reflecting lower switch cost. Nonetheless, the observed BEVoCI effects and the ANOVA results highlight the need for further investigation.

**Figure 5**

*Experiment 1: Success and Confidence by Difference in Edges across groups.*

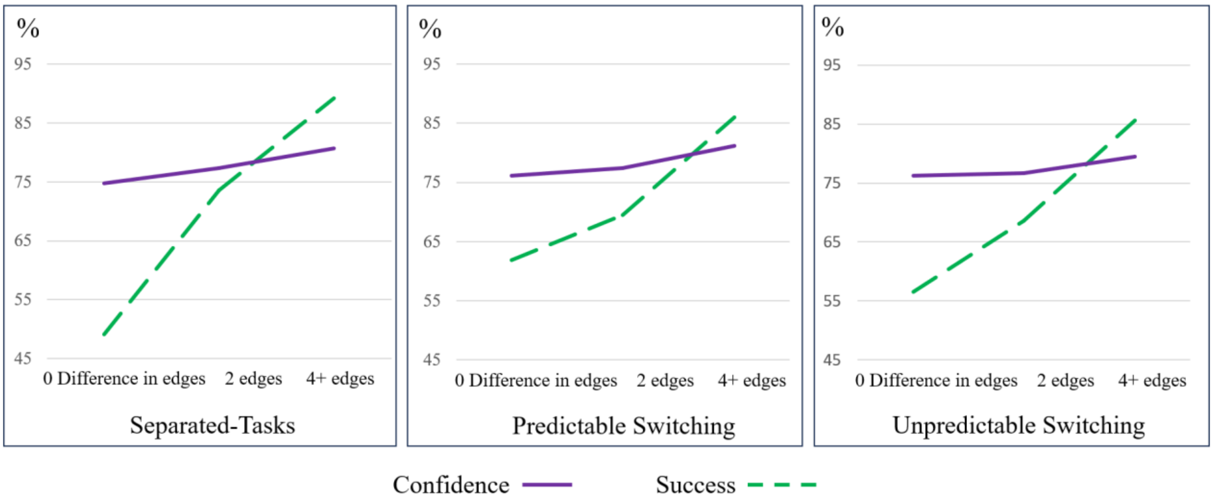


**Table 3**

*Experiment 1: One-way ANOVA between switching groups for difference in edges: success rates for no edges difference and 4+ edge difference.*

| Dependent variable | $F_{(2, 154)}$ | *MSE* | *p* | $\eta^2_p$ |
|---|---|---|---|---|
| no edge success rates | 4.41 | 334.2 | .013 | .050 |
| 4+ edge difference success rates | 0.44 | 491.2 | .644 | .005 |

As explained above, response time reflects the individual effort invested in each item. Overall, across all groups, it had a consistently insignificant effect on success rates, aligning with the baseline condition. This finding indicates that participants succeeded similarly on items regardless of whether they invested shorter or longer amounts of time. Unexpectedly, and in contrast to prior research and the baseline, confidence within the groups separated-tasks and predictable switching remained unaffected by response time (see Table 2). When examining the task-switching effect, success was not affected by the task-switching manipulation, while a significant effect on confidence observed in the unpredictable switching group, which was stronger compared to the separated-tasks group. This finding may suggest that when higher levels of cognitive load are generated, as in the unpredictable switching, the confidence declines when more time is invested in items, typically associated with greater effort and challenge.

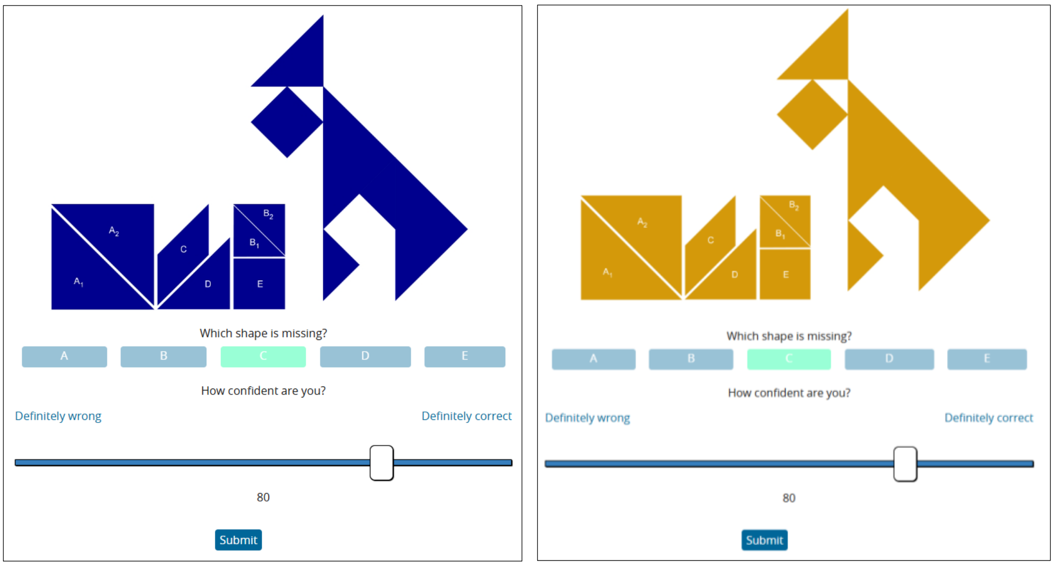
In sum, effects on success and confidence within groups (Table 2) were mostly aligned with the baseline results, while exposing that task switching moderates the effects on success as hypothesized. However, unexpectedly, confidence showed very low sensitivity to task switching. Notably, for both perimeter-area congruency and difference in edges, significant differences in effects on success rates were observed between the separated-tasks (where switching was predicted and minimal) and the unpredictable switching (where switching was random and most extreme). Interestingly, the predictable switch was not different than the other groups in perimeter-area congruency, and resembled the unpredictable switching in the sensitivity of success rates to differences in edges. Experiment 2 further examined the task-switching effect, aiming at amplifying the cognitive challenge with a secondary task being more similar to the primary task than the CRA.

## 4. Experiment 2

Switching between tasks that share common elements imposes significant switch costs (Waszak et al., 2003). This phenomenon is termed as "task-overlap" effect. Experiment 2 followed a similar design to Experiment 1, with the key difference being the use of the *Missing Tan Task* (MTT) as the secondary task, alongside the CoP task. See an example of the MTT in Figure 6. The purpose of this experiment was to minimize perceived differences between the tasks, as they both shared geometric characteristics, thereby increasing the challenge of task switching. To further emphasize this challenge and exploit the similarity between the two geometric tasks (i.e., facilitating the task-overlap effect), Experiment 2 incorporated a task-color manipulation using two color versions of the MTT: One version was presented in blue that matched the CoP task's blue color (Figure 2) - thereby generating high similarity between the tasks (i.e., task-overlap). For the second version of the MTT we used yellow stimuli, designed to resemble Experiment 1 task switching design, in which the two tasks differed in their visual appearance. The color yellow was chosen for its strong contrast with blue. According to the yellow-blue opponent mechanism, yellow has the highest luminous efficiency and spectral sensitivity (i.e., it reflects a high level of light), while blue exhibits the greatest dispersion (i.e., it is associated with more hues and tones, whereas yellow is more distinct) in the human optical system (Rabin, 2004). This combination of 'bright yellow' and 'spreading blue' creates a high visual contrast between the two colors making them easily distinguishable even for individuals with color perception abnormalities (See Opponent-process theory by Ewald Hering, in Conway et al., 2023). Consequently, the blue and yellow shapes in Experiment 2 were expected to function as distinct task characteristics.

**Figure 6**

*Experiment 2: The secondary task of the Missing Tan Task (MTT). Left Panel: Same color as the blue CoP task. Right Panel: shape color differed from the Comparison of Perimeters (CoP) task.*




Given that more consistent differences were found between separated-tasks and unpredictable switching than between the predictable switching and the other two groups, we used in this experiment only the separated-tasks and unpredictable switching groups. We aimed to examine whether the switch costs would increase when the tasks' visual similarity is more challenging (i.e., same color in addition to shapes similarity), therefore, we anticipated that the same-color condition would result in heightened switch costs, with the unpredictable switching group demonstrating the most pronounced effects. Conversely, the distinct-color condition, in which the MTT was presented in yellow, was expected to relatively reduce switch costs across both switching conditions, similarly to the CRA in Experiment 1. Thus, the separated-tasks group under the distinct-color condition was expected to experience the least switch cost.

### 4.1.1. *Method*

### 4.1.2. *Participants*

This experiment followed the same recruitment process as Experiment 1. Here, 240 participants were recruited. Following the pre-defined screening criteria as in Experiment 1 (see Participants in section 3.1.2), twelve participants were excluded: six due to low confidence variability and/or low chance of success (criteria v, vi), and six due to attention verifications or low focus (criteria i, ii, iii), leaving 228 participants that were included in the data analyses ($M_{\text{age}}$ = 36.57, *SD* = 12.76, 59% females). Applying the task switching and task color manipulations, participants were randomly assigned to the following conditions: 89 to the separated-tasks condition, 45 of whom in the same-color condition and 44 in the distinct-colors condition; 139 were assigned to the unpredictable switching condition, 68 of whom in the same-color condition and 71 in the distinct-colors condition (see Procedure).

### 4.1.3. *Materials*

The primary task was the CoP, with the same stimuli as used in Experiment 1. The secondary task was the MTT. The original Tangram task involves silhouettes formed by positioning seven geometric pieces (tans) - a square, a parallelogram, two large triangles, two small triangles, and one intermediate-size triangle. In the current study, all silhouettes were generated using six out of the seven pieces. The task was to identify the missing piece when viewing the static silhouette, without manipulating physical pieces. In the current experiment, the task included 21 items. The legend showed the seven pieces and the question “Which shape is missing?” (Figure 6) appeared on the screen throughout the task. The included silhouettes were a subset of the silhouettes used by Ackerman (2023), representing all difficulty range (20% - 100% success rates across items).

*4.1.4. Procedure*

Participants were assigned to one of two switching conditions: separated-tasks or unpredictable switching. Within each condition, participants were randomly assigned to one of two color-manipulation groups (MTT color). Participants in the separated-tasks condition completed one version of the MTT task (blue or yellow), followed by the CoP task. Participants in the unpredictable switching condition were assigned to one of the two MTT color versions, with counterbalancing applied to both the task order (CoP or MTT first) and MTT task version. Confidence ratings for the MTT was elicited by a scale ranging between 20% (chance level by a wild guess) and 100%.

*4.1.5. Results and discussion*

Descriptive results of the CoP task are presented in Table 4. Probing into the overall results, a two-way ANOVA was conducted across the condition groups (factor 1: two task-color conditions, factor 2: two task-switching groups). The results revealed the following effects (see statistical report in Table 5). (a) Success rates: task color demonstrated a significant main effect on success rates. However, task switching did not show a significant main effect on success rates, and no significant interaction effect between task-color and task-switching was found on success rates. (b) Confidence: no significant effects were found. (c) Overconfidence: task color demonstrated a significant main effect on success rates. However, task switching did not show a significant main effect on success rates, and no significant interaction effect between task color and task-switching was found on success rates. (d) Response time: no significant effects were found. (e) Resolution: While task-color and task switching did not demonstrate a significant main effect, a significant crossover interaction effect between task-color and task-switching was found for resolution. To investigate further the interaction effect on resolution, we examined the simple effects within the conditions via an independent sample t-test, revealing a significant effect between both task-color conditions, solely within the separated-tasks group, $t_{(93)} = 2.25$, $p = .026$, $d = 0.43$.

These findings align with our prediction that the distinct-color condition, associated with relatively lower cognitive load, buffers the switch costs generated by the task-switching manipulations. As presented in Table 4, this is reflected in higher success rates and lower overconfidence (indicating relatively higher calibration) across both task-switching groups, while specifically within the separated-tasks group a significant difference observed between both task-color conditions, demonstrating a higher resolution when experiencing relatively lower cognitive load (i.e., distinct-colors under separated-tasks).

**Table 4**

*Experiment 2: Means (SD) of cognitive and metacognitive measures of the Comparison of Perimeter (CoP).*

| Manipulation | Separated-tasks | | Unpredictable Switching | |
|---|---|---|---|---|
| | Same-Color | Distinct-Colors | Same-Color | Distinct-Colors |
| Dependent variables | | | | |
| (a) Success rates (%) | 66.9 (16.8) | 71.4 (14.8) | 65.0 (15.0) | 70.3 (14.6) |
| (b) Confidence (%) | 79.1 (10.8) | 76.5 (9.65) | 78.4 (9.08) | 77.4 (8.67) |
| (c) Overconfidence | 12.2 (17.8) | 4.90 (15.7) | 13.4 (15.7) | 7.11 (14.9) |
| (d) Response Time (sec.) | 16.8 (11.1) | 19.7 (15.7) | 17.6 (14.0) | 18.5 (15.7) |
| (e) Resolution (gamma) | .17 [a] (.38) | .37 [b] (.38) | .34 (.40) | .27 (.41) |

*Note*. [a,b] Significant simple effect within task color condition, $p < .05$.

**Table 5**

*Experiment 2: Two-way ANOVA significant effects of cognitive and metacognitive measures of the Comparison of Perimeter (CoP).*

| | Dependent variable | (a) Success rates (%) | | | |
|---|---|---|---|---|---|
| Effect | Manipulation | $F_{(1, 224)}$ | *MSE* | *p* | $\eta^2_p$ |
| Main | Task-color | 6.28 | 232.8 | .013 | .03 |
| | Task-switching | 0.56 | 232.8 | .454 | .002 |
| Interaction | Task-color*Task-switching | 0.02 | 232.8 | .879 | .0001 |
| | Dependent variable | (b) Confidence | | | |
| Effect | Manipulation | $F_{(1, 224)}$ | *MSE* | *p* | $\eta^2_p$ |
| Main | Task-color | 1.66 | 89.19 | .199 | .007 |
| | Task-switching | 0.00 | 89.19 | .950 | .0001 |
| Interaction | Task-color*Task-switching | 0.40 | 89.19 | .528 | .001 |
| | Dependent variable | (c) Overconfidence | | | |
| Effect | Manipulation | $F_{(1, 224)}$ | *MSE* | *p* | $\eta^2_p$ |
| Main | Task-color | 10.1 | 252.8 | .002 | .04 |
| | Task-switching | 0.57 | 252.8 | .449 | .002 |
| Interaction | Task-color*Task-switching | 0.05 | 252.8 | .818 | .0002 |
| | Dependent variable | (d) Response time | | | |
| Effect | Manipulation | $F_{(1, 224)}$ | *MSE* | *p* | $\eta^2_p$ |
| Main | Task-color | 0.71 | 208.02 | .398 | .003 |
| | Task-switching | 0.01 | 208.02 | .927 | .0 |
| Interaction | Task-color*Task-switching | 0.26 | 208.02 | .608 | .001 |
| | Dependent variable | (e) Resolution (gamma | | | |
| Effect | Manipulation | $F_{(1, 219)}$ | *MSE* | *p* | $\eta^2_p$ |
| Main | Task-color | 0.51 | 0.157 | .488 | .0002 |
| | Task-switching | 0.48 | 0.157 | .474 | .0002 |
| Interaction | Task-color*Task-switching | 5.97 | 0.157 | .015 | .03 |

Next, we applied the BEVoCI methodology as in Experiment 1. The beta coefficients and their significance are presented in Table 6.

**Table 6**

*Experiment 2: Hierarchical Multiple Regression predicting success and confidence - Beta coefficients within each group.*

| Conditions | Separated-tasks | | | | Unpredictable Switching | | | |
|---|---|---|---|---|---|---|---|---|
| | Same-color | | Distinct-Colors | | Same-Color | | Distinct-Colors | |
| | Success | Confidence | Success | Confidence | Success | Confidence | Success | Confidence |
| Perimeter-area congruency | -.02 [a] | .05 | .10 ** [b] | .04 | .10 *** [b] | .06 * | .09 *** [ab] | .01 |
| Basic shape area | -.13 *** | -.17 *** | -.06 | -.17 *** | -.07 * | -.21 *** | -.08 ** | -.18 *** |
| Difference in edges | .21 *** | .10*** | .17 *** | .09 *** | .15 *** | .07 ** | .23 *** | .08 *** |
| Response time | .03 | -.03 | -.01 | -.11 ** | .04 | -.03 | .02 | .02 |

*Note*. Significance of a cue as a predictor, * $p < .05$, ** $p \leq .01$, *** $p \leq .001$

Gray font: Insignificant betta coefficient.

[a,b] Different letters signify a significant pairwise difference between switching conditions in success rates or confidence, $p \leq .05$.

**Table 7**

*Experiment 2: Statistical report of the hierarchical multiple regression exposing exclusive interactive effects between task switching and stimuli color in perimeter-area congruencies as a predictor of success rates.*

| Manipulation | Main effects | | Interaction | Simple effects | |
|---|---|---|---|---|---|
| Dependent Variable | Task-color | Task-switching | | Task-color | Task-switching |
| Success rates | $t = 2.36$, $p = .017$ | $t = 2.60$, $p = .009$ | $t = 1.98$, $p = .046$ | $t_{(1711)} = 2.37$, $p = .017$ | $t_{(2203)} = 2.54$, $p = .011$ |
| Confidence | $t = 0.22$, $p = .823$ | $t = 0.24$, $p = .803$ | $t = 0.83$, $p = .403$ | - | - |

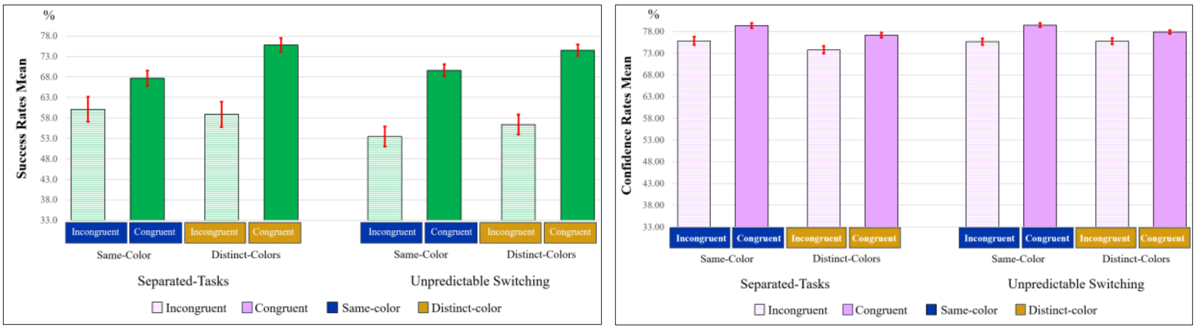
*Note*. *DF* for main effects and interaction = 4414.

Looking at Table 6, the BEVoCI method replicated Experiment 1 in the overall pattern of cue validity and cue utilization, with somewhat more consistent effects of perimeter congruency on success. As for the effects of the manipulations, most cues revealed no significant effects (all *p*s > .06) on both success and confidence. However, perimeter area congruency showed significant main effects and an interactive effect on success rates (see statistical report in Table 7).

Overall, a significant effect of perimeter-area congruency on success rates was observed within the unpredictable switching group, replicating Experiment 1. Delving into the findings revealed that the same color with separated tasks showed a weaker effect than all other groups. For confidence, a significant effect was observed only within the unpredictable switching group, combined with the same-color condition, and it was weak. Although aligning with our prediction, this finding was unexpected, as confidence remained largely unaffected by congruency differences in Experiment 1. When examining the task-color and the task-switching effects, as presented in Table 7, perimeter-area congruency demonstrated a significant main effect of task-color and task-switching on success rates. Additionally, results revealed a significant interaction effect between task-color and task-switching on success rates. To probe into the interaction effect on success rates, we examined the simple effects of task-color and task switching. Results revealed a significant simple effect for task-switching, within the same-color condition - where the unpredictable switching group demonstrated a larger gap in success rates between congruent and incongruent pairs compared to the separated-tasks group. Confidence remained robustly blind by both task-manipulations, consistent with the findings of Experiment 1. As can be seen in Figure 7, the difference in the success beta coefficients (Table 6) in the unpredictable switching group stemmed from the decreased success rates for incongruent pairs in the same-color condition. These findings align with our prediction that the same-color condition, due to the task-overlap (i.e., tasks with cues presented in the same color) would generate more pronounced switch-costs in the unpredictable switching group. This suggests that under relatively higher cognitive load - associated with unpredictable switching and task overlap - success rates decline for challenging stimuli (incongruent pairs). In sum, this pattern replicates the relationship between task switching, congruency, and success rates observed in Experiment 1, while Experiment 2 highlights the added effect of task color, emphasizing that switch costs are relatively increased under higher cognitive load.

**Figure 7**

*Experiment 2: Differences between groups in Perimeter-area congruency on Success and Confidence. Left Panel: success results, Right Panel: confidence.*




*Note*: Error bars represent ± 1 *SE*.

Difference in edges, overall, demonstrated across all the conditions a significant effect on both success rates and confidence, robustly replicating Experiment 1 (although the weak effects). Although the interactive effect was not significant (Table 6), for completeness in accordance with Experiment 1, we further investigate the effects of the task manipulations on success rates for shapes with no edge difference and for shapes with 4+ edge difference using a two-way ANOVA across the groups (Factor 1: two task-color conditions, Factor 2: two task-switching groups). This analysis revealed a significant main effect of task-color on success for 4+ edge difference shapes. However, the main effect of task-switching on success for 4+ edges shapes, as well as the interaction effect, were found insignificant. For success rates of no edge difference shapes, no significant effects were found (see Table 8). These findings indicate that above and beyond the task-switching groups, there is a significant gap in success rates of shapes with 4+ edge difference between the same-color and distinct-colors conditions. Figure 8 (left panel) illustrates that the increased success rates for easier shapes was observed in the distinct-color condition, compared to the same-color condition, across both task-switching groups. This finding suggests that when tasks are presented in distinct colors, involving some task-switching settings, the switch-cost is less pronounced, compared to conditions involving task-overlap (i.e., same-color). While these results align with our prediction regarding task-color effects, they differ from Experiment 1, which showed significant differences between task-switching groups in the relationship between edge differences and success rates. Nonetheless, the results offer valuable insights into predicting cognitive performance (i.e., success rates) across task-switching designs with varying visual stimuli (i.e., color). As for confidence, we did not further examine the task-manipulations effects in the ANOVA, since Figure 8 (right panel) illustrates a consistent and very similar pattern between groups.

**Figure 8**

*Experiment 2: Success and Confidence by Difference in Edges across groups.*

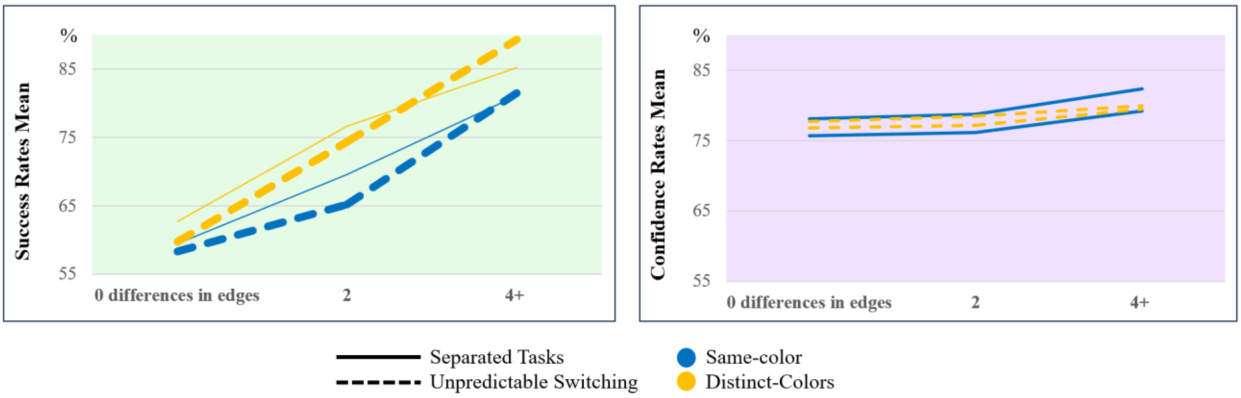


**Table 8**

*Experiment 1: One-way ANOVA between switching groups for difference in edges: success rates for no edge difference and 4+ edge difference.*

| | Dependent variable | Success rates for no edge difference | | | |
|---|---|---|---|---|---|
| Effect | Manipulation | $F_{(1, 224)}$ | *MSE* | *p* | $\eta^2_p$ |
| Main | Task-color | 0.55 | 440.1 | .453 | .002 |
| | Task-switching | 0.56 | 440.1 | .456 | .002 |
| Interaction | Task-color*Task-switching | 0.09 | 440.1 | .756 | .0004 |
| | **Dependent variable** | **Success rates for 4+ edge difference** | | | |
| Effect | Manipulation | $F_{(1, 224)}$ | *MSE* | *p* | $\eta^2_p$ |
| Main | Task-color | 5.04 | 470.3 | .025 | .02 |
| | Task-switching | 0.60 | 470.3 | .438 | .002 |
| Interaction | Task-color*Task-switching | 0.09 | 470.3 | .756 | .001 |

In sum, in Experiment 2 the effects for success and confidence within groups (Table 6) were mostly aligned with Experiment 1, while exposing that task switching combined with task-color moderates the effects on success, as hypothesized. Consistent with Experiment 1, confidence remained insensitive to task switching, as well as task-color manipulations. In general, Experiment 2 findings suggest that the relationship between heuristic-cues and success rate is not unequivocal moderated by task-switching manipulation, but may also be affected by similarity between the tasks.

## 5. General discussion

When answering knowledge questions in various settings, confidence judgments reflect our self-assessed likelihood of an answer being correct, which in turn guide our decision making regarding next steps (Ackerman & Thompson, 2017). However, when confidence is blind to task-design variations, as demonstrated in the present study, we may overlook weaknesses in our reasoning, potentially leading to suboptimal problem-solving strategies. The present study's findings contribute to the literature by emphasizing the significance of task design in optimizing cognitive performance and identifying weaknesses in reasoning processes, such as blind confidence—where confidence remains insensitive under certain conditions. Task-switching paradigms are valuable for studying cognitive flexibility and goal-directed behavior in multi-task contexts (Vandierendonck et al., 2010).

The present study aimed to investigate how task switching affects metacognitive monitoring and success in a task. Using the BEVoCI method adapted from Ackerman (2023), we delved into the task characteristics as well as task performance fluency that hypothesized to predict success and confidence. By this analysis, we aimed to determine which switching condition allows participants to perform more effectively (i.e., with minimal switch-cost) and to shed light on the underlying metacognitive processes involved. Specifically, we investigated cue utilization across three task-switching conditions: separated-tasks, predictable switching, and unpredictable switching. We examined which condition leads to greater switch cost, reflected in lower success rates and more consistent confidence levels. Additionally, we explored how task-based cues, response time, and color similarity manipulation (introduced in Experiment 2) moderated the success and confidence, with a particular focus on differences in cue validity and cue utilization across the switching conditions.

### 5.1. Effects of task switching on the predictive power of task characteristics

As anticipated, the effects of task-switching settings depended on the heuristic-cue type. Specifically, success was sensitive to the task-switching manipulation depending on the task-based cues. The strongest sensitivity to the perimeter-area congruency was in the unpredictable switch condition, while the strongest sensitivity to difference in edges was in the separated tasks condition. Basic shape area and response time did not show success prediction differences among the conditions. Confidence was blind to condition changes in both perimeter-area congruency and basic shape area. Its sensitivity to difference in edges was in line with the strongest predictive value for success in the separated tasks condition. Overall, Experiment 1 reveals a complex pattern of results that clearly demonstrates sensitivity to the task switch manipulation, but not a straight forward one. Experiment 2 aimed to shed more light on this complex pattern of results by challenging similarity between the switched tasks more than in Experiment 1. Using two geometric tasks and adding the task-color manipulation revealed some similarities and some differences from Experiment 1. Both success and confidence were hardly sensitive to the manipulations. Only success showed an effect of perimeter-area congruency (see Table 6 and Table 7). Overall, across both experiments, success was more sensitive to task manipulations (switching and color), while confidence remained largely unaffected, highlighting a dissociation between effects on success and confidence. These differential effects lead us to consider dissociations between effects on success and on confidence.

### 5.2. Heuristic cues' effects

As presented above, the results showed different patterns of the task-switching effect on success between the perimeter-area congruency and difference in edges cues. These sensitivity differences may reflect cue-specific processes that go beyond the traditional switch costs. For understanding the interaction between task-switching and task-based cues, we suggest adapting theoretical concepts from classic attention research of global and local visual processing. *Global processing* involves a holistic perception of a shape as a whole, while *local processing* involves detailed analysis of figure components (Navon, 1977). Overall, global processing has been found to have priority over local processing and thus to be processed faster (Gerlach, 2020; Martin, 1979; see also Navon's *Nested Letter Identification Task*; Navon, 1977). In analogy, we suggest that the task-based cues derived from Ackerman's (2023) work may lead to varied success and confidence patterns depending on the cue type and the task-switching condition. That is, task switching condition may have a differential effect when depending on global and local cues. Specifically, we see the perimeter-area congruency cue as requiring global processing, as it perceived as a cohesive whole, and relies on the overall shape dimensions and contour (i.e., area and perimeter) with minimal detailed analysis. In contrast, we see the difference in edges cue as requiring local processing, as it focusing on smaller components of the edge differences, which demands more attention to finer visual details within the whole shape. Although prior literature in this area has primarily focused on letters stimuli (e.g., Navon, 1977) or geometric shape stimuli based on Navon's work (e.g., Kimchi & Palmer, 1982), this theoretical division suggests a foundation for future systematic investigation, as it may offer insights into how task switching and task-based cues might relate within the observed results.

In the present study, the results revealed that when global processing was required (i.e., perimeter-area congruency), it was more challenging for the group that experienced higher cognitive load to succeed on the challenging stimuli. Conversely, when local processing was required (i.e., difference in edges), it was more challenging for the group that experienced the lowest cognitive load to succeed on the challenging stimuli. These findings may be explained by the cognitive load experienced by participants in each group, suggesting that higher cognitive load likely heightened participants' overall attentiveness, leading them to invest more effort in the task and focus more on details, thus achieving higher success when the local processing cues are in their easier form (in this case, high difference in edges). In contrast, operating under relatively lower cognitive load may have buffered the attention control, making it more difficult for participants to focus on the detailed analysis required for the local-processing shapes (e.g., focal-task engagement: Sörqvist et al., 2016).

Ackerman (2023) highlighted that each considered cue (i.e., task based) may lead to a unique result, above and beyond other cues. This conclusion suggests that each cue has its own distinct influence, depending on its characteristics (e.g., global or local processing) and the task settings. Although prior literature (Brass et al., 2003; Gopher, 2000; Kiesel et al., 2010; Rogers & Monsell, 1995) have robustly indicated that task switching results in high switch cost, other factors may also moderate the relationship between task switching and performance, as demonstrated in the present study.

Furthermore, Chiu and Egner (2017) suggested that a stimulus can signal not only a specific task but also the need to switch or maintain the current task by directing cognitive resource reconfiguration. Logan and Bundesen (2003) argued that participants do not explicitly switch tasks de facto, but rather perceive the overall task as unified, with stimulus characteristics signaling the need for a switch.

Within this framework, goal changes are triggered by embedded cues that guide responses, emphasizing the importance of stimulus design in task-switching and the critical role heuristic cues play in everyday tasks. For example, in a user interface, dynamic changes in visual cues—such as color-coded icons or highlighted buttons—signal when users should switch between tasks. These embedded cues, like transitioning from a 'draw' tool (global processing) to an 'edit' tool (local processing), guide responses and influence performance based on the nature of the cue. We suggest that, just as perimeter-area congruency requires global processing, product design can prioritize certain cognitive processes depending on the task, demonstrating how stimulus characteristics can drive task-switching effects (i.e., switch costs) within a unified workflow. This concept may help product designers create intuitive user interfaces by leveraging embedded cues to guide users through tasks without explicit instructions. At the very least, it can shed light on user behavior and cognitive load pain points, which, in turn, can enhance task performance, usability, and ultimately lead to a more successful user experience.

### 5.3. Disassociation between the effects of response time on confidence and success

As presented, success was robustly blind to response time effect, regardless of the task-manipulation, while confidence, although the weak effects, showed relatively slightly more sensitivity to response time. Interestingly, only the unpredictable switching condition replicated previous findings of confidence sensitivity to response time. This is interesting because this is the group we aimed to put under the highest cognitive load. Nevertheless, overall, confidence showed minimal sensitivity to response times, contradicting expectations under neutral conditions (Baseline condition; Ackerman, 2023). This finding in not consistent with prior literature, which has consistently found that confidence is oversensitive to response time, a phenomenon well-established in metacognitive research (e.g., Ackerman & Zalmanov, 2012).

Studies has been emphasizing a potential link between confidence and response time, whether it’s within an item or whether it is between challenging and easier items as in the present study. For example, Shynkaruk and Thompson (2006) have shown that confidence judgments tend to increase with additional time to process information within an item, while Ackerman (2014) have demonstrated that time is negatively correlated with confidence judgement, with low judgments after effortful processing (i.e., challenging items). In general, these response time patterns imply that time investment and fluency may interact with metacognitive assessments, revealing how mental resources influence confidence judgments independently of task performance. However, despite response time reflecting cognitive effort, confidence still seemed "blind" to these variations in the present study. This unexpected pattern suggests that confidence judgments might be less associated by response time than often claimed in the literature, or that this association depends on other factors, such as the task type. For instance, Ackerman and Hagar (2012) demonstrated that in one task (e.g., CRA) resulted with strong negative correlation between confidence and response time, while in another task (e.g., CRT) this relationship was negative and weak.

These inconsistencies in time-confidence relationships may be supported through *Cognitive Load Theory* (CLT) which suggests that mental effort - induced by task design and processing demands - affects task performance (Scheiter et al., 2020). However, Scheiter et al. (2020) found a stronger sensitivity of confidence to response time than that of success. Interestingly, they found weaker sensitivity of other judgments, of difficulty and of effort, to response time. As our switching manipulations were intended to affect the cognitive load demands, the present study's findings contribute to the CLT by revealing conditions of insensitivity of confidence factors that affect success, and thus generating mismatches between the two. These findings call for further research on how different types of cognitive load (e.g., mental effort vs. task-switching) interact with metacognitive processes and allow predicting conditions particularly prone to biases.

More broadly, the combined findings of the task-based cues expose that confidence was sensitive to some cues under some conditions but not for others, and most importantly, not in line with effects on success, suggesting a double dissociation between the two (Ackerman, 2023). This dissociation supports a nuanced view of metacognitive processing, where confidence judgments operate under distinct cognitive mechanisms compared to task performance metrics. Prior studies proposed that metacognitive judgements are affected by people beliefs (Bjork et al., 2013; Frank & Kuhlmann, 2017). For example, Shynkaruk and

Thompson (2006) demonstrated that confidence is influenced by factors external to the task, such as participants' beliefs about the relationship between time and accuracy (i.e., how response time affects the link between success and metacognitive judgments). Similarly, Mueller and Dunlosky (2017) demonstrated the impact of beliefs about stimulus' fluency. In their study, participants were led to believe that one color was easier to process than the other. Although color did not consistently influence final performance (e.g., success rates), participants' confidence judgments were significantly higher for words printed in the color associated with more fluent processing. Beliefs, which we did not examine, may explain why confidence in the present study was largely unaffected by task manipulations. In general, empirically identifying the mediators of effects on judgments remains a critical agenda for advancing metacognitive theory.

### 5.4. Resolution and reasoning

The results of the present study demonstrated higher resolution when experiencing relatively lower cognitive load (i.e., distinct colors under separated tasks), supporting our hypothesis that lower cognitive load reduces task-switching costs. Resolution is a key to distinguishing between what we know and what we do not. Our metacognitive judgments help us allocate time effectively across tasks, making it essential to recognize what we have

mastered versus what still needs to be processed (Thiede et al., 2003). Metacognition plays a crucial role in developing deep thinking by making individuals aware of their thinking processes, in order to improve them for better knowledge acquisition. Critical thinking relies on these metacognitive mechanisms functioning well, helping individuals understand their actions, emotions, and mistakes to correct them (Rivas et al., 2022).

However, resolution still varies across studies. Given the limited demonstrations of the effects on resolution, further investigation remains crucial. Recognizing the factors that predict resolution or situations where resolution is impaired, may help individuals improve both accuracy (i.e., success rates relative to metacognitive judgments) and effort regulation (i.e., response time). For instance, resolution is closely linked to the misleading nature of the task and may conceal differences in cues predicting success and confidence (Ackerman, 2023). Resolution is fundamentally about making decisions between two possibilities - the better option versus the less favorable one. These self-regulation strategies are a central metacognitive aspect (Rivas et al., 2022).

### 5.5. Conclusion

The present study aimed to expose factors affecting reasoners behavior, while examining the effects of task design, in terms of the switching among the tasks as well as the tasks themselves, on metacognitive judgement and actual success. More specifically, we aimed to shed new light on the metacognitive processes underlying final performance.

The results emphasize the need for further investigation into the effect of various task-switching designs on confidence and success, as well as the distinction between them, to better understand the patterns observed in the present study. The results underscore the examined heuristic cues and their effects, in conjunction with task-switching designs. Additionally, the results suggest that a switch cost may not be exclusively induced by switching between tasks per se, but could instead be generated by stimuli switching within the same task. This direction requires further investigation, as it may explain the pattern of results demonstrated in Experiment 2.

Koriat (1997) demonstrated that metacognitive judgments are highly sensitive to intrinsic cues, but less sensitive to extrinsic cues. Intrinsic cues refer to the inherent properties of the items, such as the time and effort an item requires, while extrinsic cues relate to external conditions influencing the cognitive process, such as the number of items to be remembered or serial position information (Castel, 2008; Koriat, 1997). This may explain why confidence was largely unaffected by the task-switching effect in the current study, as it is more closely associated with extrinsic cue types. Note that in the present study, response time was measured as an objective heuristic cue rather than a subjective individual experience and therefore qualifies as an extrinsic cue. However, these findings, which relate to metacognitive judgments of learning in meta-memory research, differ from the present study's focus on meta-reasoning. Unlike memory, reasoning unfolds over a longer period, which may cause the underlying processes to operate differently (Ackerman & Thompson, 2015). For instance, a prior study demonstrated opposite patterns of correlations between meta-reasoning and meta-memory judgments with metacognitive processes. Moreover, differences were detected in the cues predicting reasoning and memory performance (Ackerman & Beller, 2017). Nonetheless, there are strong analogies between meta-memory and meta-reasoning. For instance, some reasoning tasks can be solved by retrieving an answer from memory, while others require deliberate, time-intensive processes that place significant demands on working memory (see Ackerman & Thompson, 2015; Beilock & DeCaro, 2007). Thus, this distinction further emphasizes the need for further investigation into the impact of different heuristic cues on both confidence and success in reasoning, particularly in the context of task-switching.

As Ackerman (2023) noted, reliable metacognitive judgments regarding each task item are important for effective self-regulated performance. Therefore, understanding when confidence is sensitive to different task designs may significantly influence decision-making, strategy adaptation, and the validity of conclusions in reasoning.

**References**

Ackerman R. (2014). The diminishing criterion model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General, 143*(3), 1349-68

Ackerman, R. (2023). Bird's-Eye View of Cue Integration: Exposing Instructional and Task Design Factors Which Bias Problem Solvers. *Educational Psychology Review, 35*(55), 1-46.

Ackerman, R., & Beller, Y. (2017). Shared and distinct cue utilization for metacognitive judgements during reasoning and memorization. *Thinking & Reasoning, 23*(4), 376-408.

Ackerman, R., & Goldsmith, M. (2011). Metacognitive Regulation of Text Learning: On Screen Versus on Pape. Journal of Experimental Psychology: Applied, 17(1), 18 –32.

Ackerman, R., & Zalmanov, H. (2012). The persistence of the fluency–confidence association in problem solving. *Psychonomic Bulletin & Review, 19,* 1187–1192.

Ackerman, R., Parush, A., Nassar, F., & Shtub, A. (2016). Metacognition and system usability: Incorporating metacognitive research paradigm into usability testing. *Computers in Human Behavior, 54*, 101-113.

Ackerman, R. & Thompson, V. (2017). Meta-reasoning: monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences 21*(8), 1-21.

Ackerman, R., Yom-Tov, E., & Torgovitsky, I. (2020). Chapter four - Self-Regulated Learning and the Allocation of Study Time. *Cognition, 199*, 1-13.

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, & Computers, 35,* 634–639.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology, 64*(1), 417-444.

Brass, M., Ruge, H., Meiran, N., Rubin, O., Koch, I., Zysset, S., Prinz, W., von Cramon, D. Y. (2003). When the same response has different meanings: recoding the response meaning in the lateral prefrontal cortex. *NeuroImage, 20,* 1026-1031.

Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory and Cognition, 36,* 429-437.

Chiu, Y. C., & Egner, T. (2017). Cueing cognitive flexibility: Item-specific learning of switch readiness. *Journal of Experimental Psychology: Human Perception and Performance, 43*(12), 1950–1960.

Fiedler, K., Ackerman, R., & Scarampi, C. (2019). Metacognition: monitoring and controlling one's own knowledge, reasoning, and decisions. In R. J. Sternberg & J. Funke (Eds.), *The Psychology of Human Thought: An Introduction* (pp. 89-111). Heidelberg University Publishing.

Finley, J. R., Benjamin, A. S., & McCarley, J. S. (2014). Metacognition of multitasking: How well do we predict the costs of divided attention? *Journal of Experimental Psychology: Applied, 20*(2), 158–165.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*(443), 1-9.

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to Vol. effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(5), 680–693.

Gerlach Ni, C. & Poirel, N. (2020). Who's got the global advantage? Visual field differences in processing of global and local shape. *Cognition, 195*(5), 104-131.

Gonzalez-Gomez, B., Nunez-Pena, M. I., & Colome, A. (2023). Math anxiety and the shifting function: An event-related potential study of arithmetic task switching. *European Journal of Neuroscience, 57*(11), 1848-1869.

Gopher, D. (1994). Analysis and measurement of mental workload. In G. d'Ydewalle, P. Eelen, & P. Bertelson (Eds.), *II. International perspectives on cognitive sciences* (pp. 265–291). London: Lawrence Erlbaum. Chap. 13.

Gopher, D., Ben-Eliezer, D., Levine, A. (2022). Imposed load versus voluntary investment: Executive control and attention management in dual-task performance. *Acta Psychologica, 227*, 1-13.

Hazeltine, E. (2024). What are we measuring when we measure task switch costs? Current Opinion in *Behavioral Science, 56,* Article 101352. https://doi.org/10.1016/j.cobeha.2024.101352

Hazeltine, E., Ruthruff, E., & Remington, R. W. (2006). The role of input and output modality pairings in dual-task performance: Evidence for content-dependent central interference. *Cognitive Psychology, 52*(4), 291–345.

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology, 14, 89,* 81.

Koch, P. E. (200310. The role of external cues for endogenous advance reconfiguration in task switching. *Psychonomic Bulletin & Review, 10*, 488-492.

Koch, P. E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin, 144*(6), 557.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370.

Lauterman, T., & Ackerman, R. (2014). Overcoming screen inferiority in learning and calibration. Computers in Human Behavior 35(1), 1-37.

Lauterman, T., & Ackerman, R. (2023). Initial Judgment of Solvability: Integrating heuristic cues with prior expectations regarding the task. *Thinking & Reasoning,* DIO: 10.1080/13546783.2023.2214378

Logan, G. D., & Bundesen, C. (2003). Clever homunculus: Is there an endogenous act of control in the explicit task-cuing procedure? Journal of Experimental Psychology: Human Perception and *Performance, 29*(3), 575–599.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*(3), 134-140.

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language, 93,* 245–258

Nelson, T., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. Psychology of Learning and motivation, 26, 125–322.

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.

Rabin, J. (2004). Quantification of color vision with cone contrast sensitivity. *Vision Neuroscience, 21,* 483–485.

Raven, J., & Raven, J. (2003). Raven Progressive Matrices. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp. 223–237). Kluwer Academic/Plenum Publisher.

Rivas, S. F., Saiz, C., & Ossa, C. (2022) Metacognitive Strategies and Development of Critical Thinking in Higher Education. *Frontiers in Psychology, 13,* 1-13.

Rogers, R., & Monsell, S. (1995). The costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: Human Perception and Performance, 124,* 207-231.

Sannomiya, M. & Ohtani, K. (2015). Does a dual- task selectively inhibit the metacognitive activities in text revision? *Thinking Skills and Creativity, 17*, 25-32.

Scheiter, K., Ackerman, R., & Hoogerheide, V. (2020). Looking at Mental Effort Appraisals through a Metacognitive Lens: Are they Biased? Educational Psychology Review, 32, 1003-1027. Stavy, R., & Babai, R. (2008). Complexity of shapes and quantitative reasoning in geometry. *Mind, Brain, and Education, 2*(4), 170-176.

Shekhar, M., & Rahnev, D. (2021). Sources of Metacognitive Inefficiency. *Trends of Cognitive Science, 25*(1), 12-23.

Sörqvist, P., Dahlström, Ö., Karlsson, T., & Rönnberg, J. (2016). Concentration: the neural underpinnings of how cognitive load shields against distraction. *Frontiers in Human Neuroscience, 10*(221), 1-10.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73.

Thompson, V. A., Prowse Turner, J. A., Pennycook, G., Ball, L., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128*, 237-251.

Tobler, S., & Kapur, M. (2023). Metacognitive calibration: a methodological expansion and empirical application. *International Conference of the Learning Sciences, Montreal, Canada,* 1-9.

Unkelbach, C. (2007). Reversing the Truth Effect: Learning the Interpretation of Processing Fluency in Judgments of Truth. *Journal of Experimental Psychology Learning Memory and Cognition, 33*(1), 219-230.

Vandierendonck, A., Liefooghe, B., & Liefooghe, F. (2010). Task Switching: Interplay of Reconfiguration and Interference Control. *Psychological Bulletin, 136*(4), 601–626.

Vuorre, V., & Metcalfe, J. (2022). Measures of relative metacognitive accuracy are confounded with task performance in tasks that permit guessing. *Metacognition and Learning, 17,* 269-291