REVISTA DE ADMINISTRAÇÃO DA UFSM
Brazilian Journal of Management
ReA UFSM
ISSN 1983-4659
OPEN ACCESS
Rev. Adm. UFSM, Santa Maria, v. 18, n. 2, e1, 2025 https://doi.org/10.5902/1983465989667
Submitted: 11/06/2024 • Approved: 05/21/2025 • Published: 06/27/2025

**Original Article**

# The influence of task complexity factors on decision-making performance

## O impacto de fatores da complexidade da tarefa na performance na tomada de decisão

**Carolina Schneider Bender**[I], **Mauri Leodir Löbler**[I], **Eliete dos Reis Lehnhart**[I]

[I] Universidade Federal de Santa Maria, Santa Maria, RS, Brazil

## ABSTRACT

**Objective:** This study aims to understand the impact of complexity factors on human decision-making performance in an economic problem (*Knapsack Problem*), which is analogous to numerous situations individuals encounter daily related to the practices of administrators and managers.
**Methodology:** The research employed an experimental design with 41 participants, varying the number of items and two metrics of computational complexity: "input size" and "instance correlation." Performance was assessed by measuring optimization performance, relative performance, response time (RT), and confidence.
**Results:** The findings reveal a significant impact from manipulating the number of items, leading to a decrease in optimization performance and confidence, alongside an increase in RT. A phase transition was observed in relative performance, where participants managed increases from 5 to 6 items despite longer task completion times; however, this compensation was no longer feasible at 7 items. The input size measure was significantly associated with all dependent variables, explaining 51.84% of the variation in RT.
**Practical Implications:** These results provide actionable insights for managers and professionals dealing with complex decision-making tasks, showing that even small increases in the number of elements can significantly raise cognitive load, impair performance, delay responses, and reduce confidence. To mitigate these effects, decision environments - particularly those involving human-algorithm interaction - should be designed to simplify task structure, present information incrementally, and adapt dynamically to problem complexity.

**Keywords:** Complexity; Optimization problem; *Knapsack Problem*; Decision-making

## RESUMO

**Objetivo:** Este estudo tem como objetivo compreender o impacto de fatores da complexidade na performance na tomada de decisão humana em um problema econômico (*Knapsack Problem*), correlato de inúmeras situações enfrentadas diariamente pelos indivíduos relacionadas à prática de administradores e gestores.

**Metodologia:** A pesquisa empregou um delineamento experimental com 41 participantes, variando o número de itens e duas métricas de complexidade computacional, "*input size*" e "i*nstance correlation*". Foi avaliada a performance por meio da mensuração do desempenho de otimização, desempenho relativo, RT e confiança.

**Resultados:** Os resultados evidenciam um impacto robusto da manipulação do número de itens, resultando na redução do desempenho de otimização e confiança e aumento do RT. Encontrou-se uma transição de fase para o desempenho relativo, na qual os sujeitos suportaram aumentos de 5 para 6 itens em detrimento de realizar a tarefa em maior tempo, no entanto, para 7 itens tal compensação não foi mais possível. A medida *input size* associou-se significativamente com todas as variáveis dependentes, sendo capaz de explicar 51,84% da variação RT.

**Implicações práticas:** Os resultados fornecem insights práticos para gestores e profissionais que lidam com tarefas complexas de tomada de decisão, mostrando que mesmo pequenos aumentos no número de elementos podem aumentar significativamente a carga cognitiva, prejudicar o desempenho, atrasar as respostas e reduzir a confiança. Para mitigar esses efeitos, os ambientes de decisão - especialmente aqueles que envolvem interação com máquinas e softwares - devem ser projetados para simplificar a estrutura das tarefas, apresentar as informações de forma incremental e se adaptar dinamicamente à complexidade do problema.


**Palavras-Chave:** Complexidade; Problema de otimização; *Knapsack Problem*; Tomada de decisão

## 1 INTRODUCTION

Decision-making is, in itself, a complex activity that requires the agent to integrate and combine information – often from multiple sources and with high degrees of uncertainty – and to act in pursuit of an objective, whether explicit or implicit. As these internal processes are automatically carried out throughout life as a result from the human capacity of learning to adapt their behavior according to past experience, many of the decision-making situations that individuals face turns out to be resolved without a great degree of difficulty, and in some cases without even being noticed.

This dynamic arises across a wide range of situations. It is present in low-level cognitive activities, such as the motor action of picking up a glass of water, which requires perceiving the distance to be covered, as well as calibrating speed and force. It also emerges in value-based preferential decisions, such as choosing between chocolate or fruit for a snack. Moreover, it plays a role in more complex tasks, including learning, memorizing, and recalling the sequence of steps required to make coffee, drive a car, or send an email with a report attached. While these activities may pose significant challenges to a naive agent, they become routine and cognitively undemanding with repeated practice.

However, certain situations stand out in which the complexity inherent in the problem or environment significantly impacts individuals—affecting information processing, perceived difficulty, intrinsic motivation, and performance. These effects can also manifest post-decision, in the form of regret, reduced confidence in the absence of feedback, stress, and procrastination. In response to such findings and in an effort to deepen the understanding of human behavior, researchers began investigating how complexity influences cognitive processes, including information processing, strategy selection for problem-solving, and the final decision or choice (Campbell, 1988; Payne, 1976; Payne et al., 1988, 1991, 1993; Tversky & Kahneman, 1981). Empirical evidence from these studies shows that complexity affects decision time, depth of search, consistency, performance, preferences, motivation, satisfaction, cognitive load, and other factors that shape the relationship between the task and the decision-making agent.

It is important to note that decision-making complexity is not a unitary construct but a multifaceted one (Adomavicius et al., 2020), shaped by a variety of factors that depend on the domain, the theoretical paradigm adopted, and the specific characteristics of the problem under investigation. To delineate the scope of the broader project to which this study belongs, a systematic review on complexity and decision-making was conducted in 2021 by the same authors of the present research. As a result, this study adopts a dual perspective: one grounded in the task complexity literature (Campbell, 1988; Liu & Li, 2012; Mackinnon & Wearing, 1980; Wood, 1986), and another based on formal mathematical structures from computational complexity theory, previously applied to examine how individuals respond to computationally difficult problems (Carruthers, 2015; Franco et al., 2021; Hidalgo-Herrero et al., 2013;

Murawski & Bossaerts, 2016). The objective of this study, therefore, is to understand the impact of complexity factors on decision-making performance.

Distinct task characteristics have been identified in the literature as contributing to perceived or objective complexity. For instance, Liu and Li (2012) proposed a framework encompassing 18 factors, including the number of elements involved in a task and the degree of variation and heterogeneity among them. In computational complexity theory, complexity is generally analyzed in terms of the time and space resources required by an algorithm to solve a problem, depending on the strategy used. Key considerations include the relationship between input size and optimization constraints, as well as the structural similarity among alternatives.

This study addresses a theoretical gap by integrating distinct approaches to complexity, which often overlap in terms of characteristics and analytical variables, yet had not been jointly examined to enable a more comprehensive understanding of their potential effects on individuals. These approaches share the foundational assumption of bounded human cognition.

By bridging research traditions that focus on aspects such as quantity, diversity, and redundancy with measures from Computational Complexity Theory—which capture task difficulty based on instance-specific features and the strategy employed—this study advances a more unified perspective. Moreover, it contributes to identifying which objective task features drive complexity, thereby enabling the design of decision-making tasks that minimize unnecessary cognitive load and reduce the likelihood of decision errors.

## 2 THEORETICAL BACKGROUND

### 2.1 Complexity in problem-solving and decision-making

Problem-solving complexity, from a reductionist philosophical perspective, can be characterized in terms of a system's hierarchical level, residing in the large number of components and their multiple interactions (Simon, 1962). Typical attributes of a complex system include the number of variables involved, the degree of connectivity and mutual dependency among these variables, the dynamic nature of the situation, the lack of clarity regarding the variables and their values, and the presence of multiple conflicting goals across different levels of analysis (Dörner & Funke, 2017).

In an effort to clarify the resources or components that make a task complex for individuals, a mapping was conducted of studies that investigated the factors contributing to the complexity of a task or decision problem (see Table 1).

Table 1 – Factors determining complexity in cognitive science and decision making

| Author | Complexity components |
| --- | --- |
| SIMON (1962) | • number of elements or parts in a system;<br>• degree and nature of the interaction between them. |
| MACKINNON; WEARING (1980) | • number of elements or parts in a system;<br>• degree to which there are interconnections between these parts and the pattern of these interconnections;<br>• degree to which uncertainties affect the system behavior. |
| WOOD (1986) | • task result;<br>• necessary information;<br>• actions required to solve the task. |
| CAMPBELL (1988) | • multiple paths;<br>• multiple results;<br>• interdependence between paths;<br>• conflict;<br>• uncertainty. |
| FUNKE (2012) | • number of variables involved;<br>• connectivity and dependence between the variables involved;<br>• temporal and developmental situation within the system;<br>• uncertainty regarding variables and values;<br>• conflicting objectives at different levels of analysis. |

Source: Prepared by the authors based on the cited studies

Thus, we identified that the most basic internal characteristic discussed as an enhancing task complexity is the number of elements (Campbell, 1988; Liu & Li, 2012; Mackinnon & Wearing, 1980; Simon, 1962; Wood, 1986). Increasing the number of items will increase the task complexity, which will affect decision-making performance, since more cognitive effort will be required given the increase in workload (Galy et al., 2012; Klepsch & Seufert, 2020; Mangion, 2017; Zagermann et al., 2016).

In terms of performance, the effects of problem complexity on individuals can be assessed in two ways. One approach evaluates the maximization of the objective function—that is, whether the optimal solution was found—referred to in this study as optimization performance (Franco et al., 2022). The other approach assesses decision quality, recognizing that finding the optimal solution is often unfeasible due to human cognitive limitations.

In this case, relative performance is considered, which measures how closely the submitted solution approximates the optimal one (Franco et al., 2022; Murawski & Bossaerts, 2016). In this context, where an "instance" refers to a single unit of a problem, the following hypotheses are proposed:

*H1: Instances with a larger number of items result in lower optimization performance.*

*H2: Instances with a larger number of items result in lower relative performance.*

Additionally, a positive relationship is suggested between task complexity and the time required to complete it, due to the increased cognitive load imposed. Increasing the number of elements in a choice task raises the cognitive effort demanded—that is, the load on working memory required to process the items and their interactions (Sweller, 1988; Sweller, Van Merriënboer & Paas, 2019). As a result, response time (RT) tends to increase, since the processes of information acquisition and alternative evaluation become more time-consuming. Accordingly, the following hypothesis is proposed:

*H3: Instances with a greater number of items result in greater RT.*

An increase in task complexity, resulting from a greater number of elements, not only requires more cognitive resources from individuals and increases the likelihood of errors but also introduces uncertainty during the decision-making process. Since decisions are generally associated with a sense of confidence regarding their quality, even in the absence of explicit or immediate feedback (Brus et al., 2021; Gavas et al., 2018; Lempert et al., 2015), a larger number of elements and interactions complicates individuals' ability to analyze and accurately assess their own decisions (Zagermann et al., 2016).

Therefore, when faced with more complex tasks, individuals report lower levels of confidence, as the ability to monitor their own performance diminishes in the presence of a higher cognitive load imposed by the problem (Aitchison et al., 2015; Double & Birney, 2018). Therefore, the following hypothese is outlined:

*H4: Instances with a greater number of items result in lower confidence.*

## 2.2 Computational Problem-Solving in Human Decision-Making

Computational complexity, rooted in computer science, refers to the study of the time and space resources required to solve problems, as formalized in Computational Complexity Theory. Time and space are conceptualized as computational resources, enabling the evaluation of a problem's inherent difficulty (Bossaerts & Murawski, 2017). Over the past decade, this rigorous mathematical framework for classifying problem difficulty has been increasingly applied to cognitive science and decision research (Carruthers, 2015; Franco et al., 2021; Hidalgo-Herrero et al., 2013; Murawski & Bossaerts, 2016), facilitating the translation of formal measures from computer science into the study of human decision-making.

Numerous computational problems have had their complexity formally characterized in the computer science literature. Among those explored within cognitive science, notable examples include the Traveling Salesman Problem (Carruthers, 2015; Hidalgo-Herrero et al., 2013), the Packing Problem (Ross et al., 2019), the Vertex Cover Problem (Burns et al., 2006), and the Knapsack Problem (KP) (Franco et al., 2021; Hidalgo-Herrero et al., 2013; Meloso et al., 2009; Murawski & Bossaerts, 2016). The latter, the Knapsack Problem, is of particular relevance, as it serves as the basis for the present research.

The Knapsack Problem (KP) is a classic combinatorial optimization problem that, as noted by Franco et al. (2021), is pervasive in everyday decision-making. The objective function of the Knapsack Problem (KP) involves selecting, from a set of items - each defined by a weight and a value - a subset that maximizes total value without exceeding a given capacity constraint. Based on these three defining attributes (value, weight, and capacity constraint), prior studies have examined how the strategies used by algorithms to solve KP relate to problem complexity (Franco et al., 2021; Murawski & Bossaerts, 2016). Two complexity-related measures identified in this context closely resemble those discussed in the task complexity literature: input size and instance correlation.

Input size refers to a measure that combines the number of items in a problem with the imposed constraint, where larger values indicate greater complexity (Murawski & Bossaerts, 2016). An increase in input size corresponds to a greater number of steps required to reach a solution, as the number of possible suboptimal combinations grows, thereby increasing the computational memory load. This aligns with the concept of task complexity as discussed in the psychological literature, particularly with respect to the presence of multiple paths and multiple potential outcomes - both of which are known to impair decision-making performance. Therefore, the following hypotheses are outlined:

*H5: Increasing input size results in lower optimization performance.*

*H6: Increasing input size results in lower relative performance.*

Additionally, *time complexity*, typically based on the number of items, serves as an indicator of the time an algorithm takes to solve a problem, and studies involving human participants have confirmed similar performance effects (Franco et al., 2021; Hidalgo-Herrero et al., 2013; Murawski & Bossaerts, 2016). Based on these findings, it is expected that increases in problem complexity, as operationalized by input size, will negatively affect decision-making performance. Accordingly, the following hypothese is proposed:

*H7: Increasing input size results in higher RT.*

An increase in the function of input size by constraint—reflecting a larger number of elements and potential interactions (Campbell, 1988; Liu & Li, 2012; Simon, 1962)—not only raises task difficulty but also impairs individuals' ability to accurately self-assess their performance. As task complexity grows, individuals face greater difficulty in monitoring the quality of their decisions (Wei & Wang, 2015), which contributes to a diminished sense of control over outcomes and, consequently, lower decision confidence. Therefore, the following hypotheses are outlined:

*H8: Increasing input size results in lower confidence.*

*Instance correlation*, in turn, is calculated using the Pearson correlation coefficient between the values and weights of the items in a given instance. Higher correlation strength is associated with greater problem complexity, as it increases the gap between continuous and integer solutions, results in ill-conditioned problem instances, and limits item variability—thereby making it harder to classify items according to criteria such as weight-to-capacity fit (Murawski & Bossaerts, 2016). As a result, the more highly correlated the items are, the more difficult it becomes—for both individuals and algorithms—to determine which items should be included in the optimal subset, ultimately impairing decision-making performance. It is worth noting that this measure aligns with the constructs of *variability* and *heterogeneity* as defined in the task complexity framework proposed by Liu and Li (2012). Based on this reasoning, the following hypotheses are proposed:

*H9: Higher instance correlation results in lower optimization performance.*

*H10: Higher instance correlation results in lower relative performance.*

*H11: Higher instance correlation results in higher RT.*

*H12: Higher instance correlation results in lower confidence.*

# 3 MATERIALS AND METHODS

## 3.1 Research design and operationalization

The present study employed an experimental design based on the collection of repeated measures from all participants across all conditions of the independent variables. The experiment was conducted in a laboratory setting to minimize the influence of extraneous variables, ensuring a rigorously controlled environment with regard to silence, lighting, and thermal comfort. The experimental protocol was approved by the Human Research Ethics Committee of Universitat Pompeu Fabra (CIREP Approval Number 242). All procedures complied with local ethical guidelines and regulations, and written informed consent was obtained from each participant prior to the experimental sessions.

The sample size was determined using G*Power 3.1 (Erdfelder et al., 2009). Based on the following parameters - repeated-measures ANOVA with within-subject factors in a 2x3 factorial design, a medium effect size ($f = 0.25$), significance level ($\alpha = 0.05$), and statistical power ($1–\beta = 0.90$) - the recommended sample size was 36 participants. To account for potential attrition or data exclusion, a 15% margin was added, leading to a final target of 42 participants.

The selection of participants was conducted using the CBC Lab database (Center for Brain and Cognition Laboratory at Universitat Pompeu Fabra). This database includes information from individuals who have expressed interest in participating in research, with registration being open and free of charge. A total of 498 individuals were randomly contacted via email, and individual sessions were scheduled until the target sample size was reached.

One participant (ID 040) was excluded before any data analysis took place, as they were unable to complete all 240 instances of the task due to a mouse malfunction. Since less than 50% of the total attempts were reliably completed by this participant, they were excluded to avoid any issues with the statistical analysis. The final sample consisted of 41 participants (24 self-identified women, 17 self-identified men, none identifying as another gender), aged between 19 and 42 years (92.7% were 32 years old or younger). Among them, 58.6% were students (34.1% were studying and working; 7.3% were only working).

The experiment took place in an isolated room at the CBC Lab. Upon arrival, participants were provided with key information about the lab, the COVID-19 protocol, and the task. After signing the informed consent forms, they were escorted to a room with thermal and noise insulation, equipped with a mouse and a screen where the instructions were displayed. Following an overview of the procedure, participants underwent a training phase consisting of 12 practice trials, which were excluded from the data analysis. This phase served to ensure participants understood the task. During this time, they had the opportunity to ask the experimenter any questions before proceeding to the main task. The task was entirely constructed and presented in English. A B1 proficiency level in English was required for participants to be eligible for the study.

Once the main task began, participants remained alone in the room and completed the experiment independently, with no further contact from the experimenter. Participants were informed that they could withdraw at any time. The data collection process lasted between 40 minutes and 1 hour, with sessions conducted between June 21 and July 18, 2022.

The experiment was fully developed using the MATLAB programming language (version R2020b), incorporating the Psychophysics Toolbox Version 3 (PTB) for stimulus synthesis and presentation. This setup enabled precise control over the experiment and accurate data capture. Data analysis was conducted using R software.

### 3.2 Variables

The number of items was manipulated (using sets of 5, 6, and 7 items), as this is considered a fundamental characteristic contributing to task complexity. In addition to item count, two measures derived from computational complexity theory were used to assess problem difficulty: input size and instance correlation.

The input size reflects the scale of the problem and is calculated by multiplying the number of items by the base-2 logarithm of the instance's value restriction ($n \times log_2 R$). Higher input size values are associated with greater complexity, as they account for interactions between the number of items and the magnitude of the instance constraints.

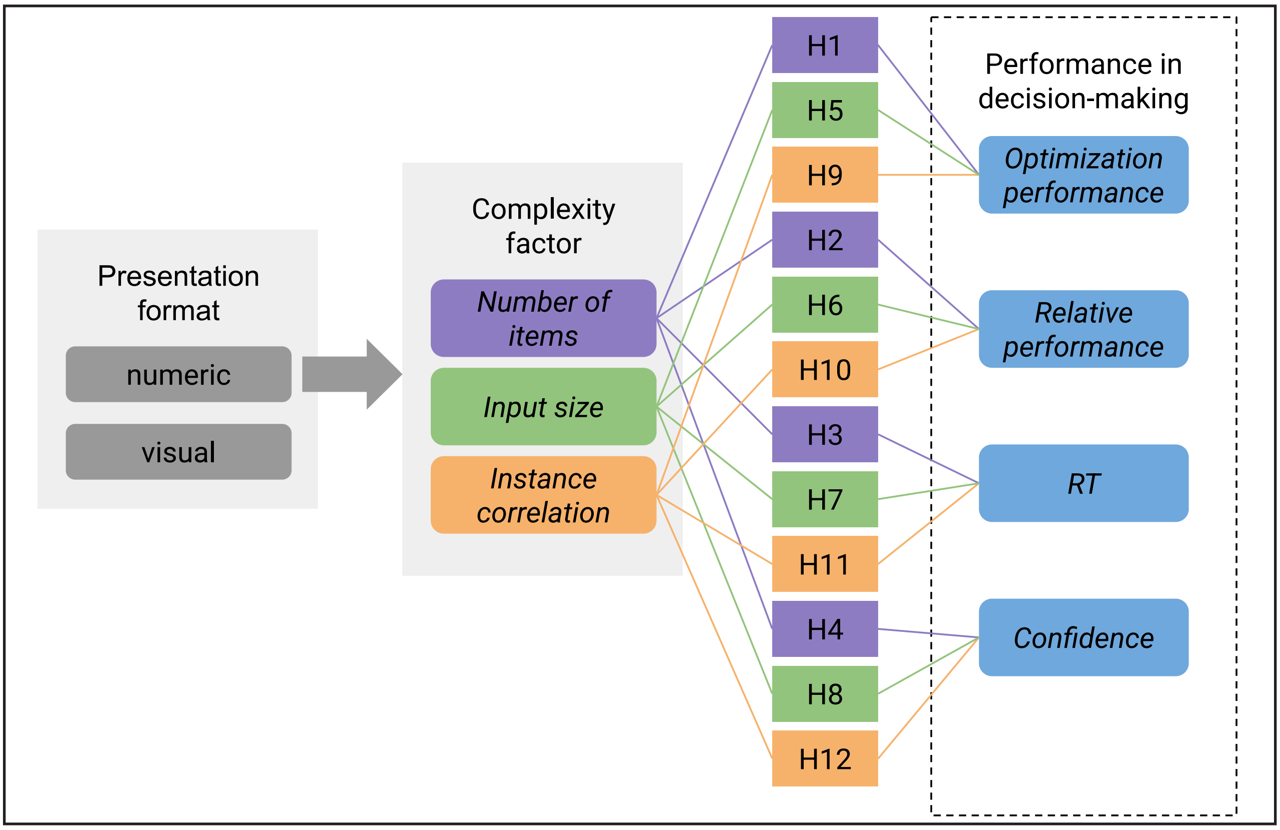
The instance correlation measure captures the relationship between item values and weights. It is expressed as the correlation coefficient ($corrcef \frac{valor}{peso}$) between these two variables. Higher absolute correlation values (i.e., closer to +1 or -1) indicate greater complexity, as they make it more difficult to determine the optimal item selection due to reduced differentiation between high-value and low-weight items.

The task was administered to participants with the information presentation format manipulated at two levels: numerical and visual. This manipulation was part of a broader research agenda aimed at exploring multiple dimensions of problem complexity. While the numerical format had been employed in prior studies (Franco et al., 2021; Murawski & Bossaerts, 2016), the visual format was newly developed and validated by the authors for the purposes of this investigation. Accordingly, the presentation and analysis of results carefully distinguish between the levels of this manipulation to ensure methodological transparency and to preserve the interpretability of the findings within the scope of the main research objective.

The dependent variables in this study comprise several indicators of decision-making performance. These include: (1) optimization performance, measured as a binary outcome indicating whether the participant identified the optimal solution (1

= correct, 0 = incorrect); (2) relative performance, defined as the proportion of the participant's response value relative to the optimal solution (ranging from 0 to 1); (3) response time (RT), calculated as the elapsed time between stimulus onset and submission of a response; and (4) self-reported confidence, assessed on a continuous scale from 0 to 1, indicating the participant's perceived certainty that the submitted response was correct. Figure 1 illustrates the research model, highlighting the relationships between the manipulated independent variables and the measured dependent outcomes. Table 2 summarizes all variables.

Figure 1 – Research Model




Source: Prepared by the authors

Hypotheses H1 through H12, declared in the previous section, aim to investigate the isolated effects of each task complexity factor on individual performance, analyzed separately for each information presentation format.

Table 2 – Overview of Manipulated and Measured Variables

| Variable | | Definition | Source |
|---|---|---|---|
| Independent | Number of items | Number of elements in each problem instance (5, 6, or 7 items), representing a basic indicator of task complexity. | Task complexity literature (Campbell, 1988; Wood, 1986; Liu & Li, 2012) |
| | Input size | Computational measure calculated as the product of the number of items and the base-2 logarithm of the value constraint. | Computational Complexity Theory (Murawski & Bossaerts, 2016) |
| | Instance correlation | Pearson correlation between item weights and values, indicating the similarity structure of the problem instance. | |
| Dependent | Optimization performance | Binary indicator of whether the submitted response matches the optimal solution (1 = correct, 0 = incorrect). | Decision performance literature (Franco et al., 2021; Murawski & Bossaerts, 2016) |
| | Relative performance | Continuous measure (0–1) indicating how close the submitted response is to the optimal solution. | |
| | Response time (RT) | Time (in seconds) from stimulus presentation to response submission. | Cognitive psychology and decision-making (Heitz, 2014; Chen, 2013) |
| | Decision confidence | Self-reported confidence level (0–1) provided by participants immediately after submitting a response, prior to feedback. | Metacognition and confidence literature (Yeung & Summerfield, 2012; Lempert et al., 2015) |

Source: Prepared by the authors

## 3.3 Experimental Task

A novel experimental task was developed for this study, based on Franco et al. (2021) and Murawski and Bossaerts (2016), grounded in the combinatorial optimization framework known as the Knapsack Problem. Specifically, the 0-1 version of the problem was employed, in which participants are required to decide whether each of the available items should be included in the “knapsack” or not. The optimization variant of the problem was selected due to its NP-hard classification, which ensures an adequate level of computational complexity for cognitive experimentation. The formal structure of the problem is defined as follows:

*Given a set of n items* $(i)$,

*each with a weight (w) =* $\{w_1, \dots, w_n\}$ *and a value* $(v_i)$ = $\{v_1, \dots, v_R\}$,

*and a total value constraint* $(R)$,

*find the subset of items that:*

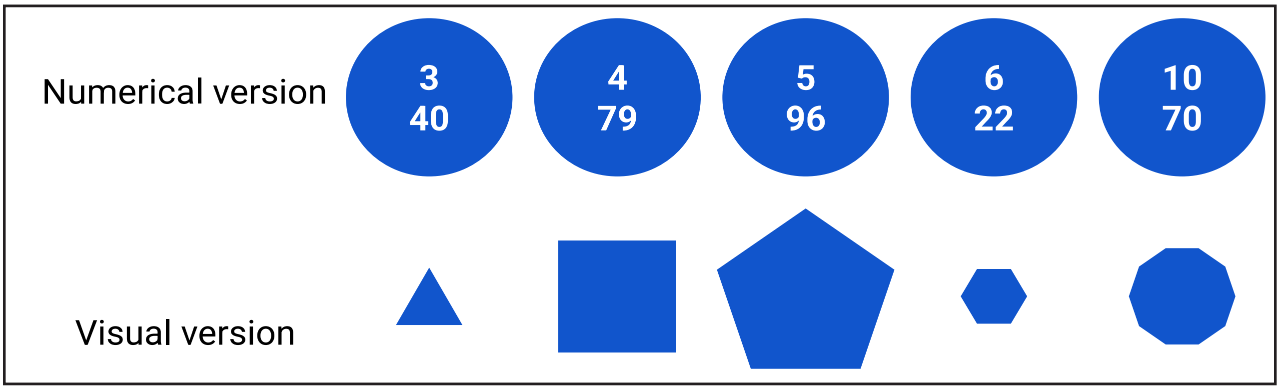
$$maximize \sum_{i=1}^{n} w_i x_i$$

$$restrict\ to \sum_{i=1}^{n} v_i x_i \leq R;$$

$$and\ x_i \in \{0,1\}, to\ i = 1, 2, \dots, n.$$

The experimental task consists of two variants (numerical and visual) of the Knapsack Problem optimization version (Figure 2). In both, for each attempt, participants were presented with an instance of the problem, formed by a set of  items, with their respective weights , values  and the value constraint.

Figure 2 – Task versions

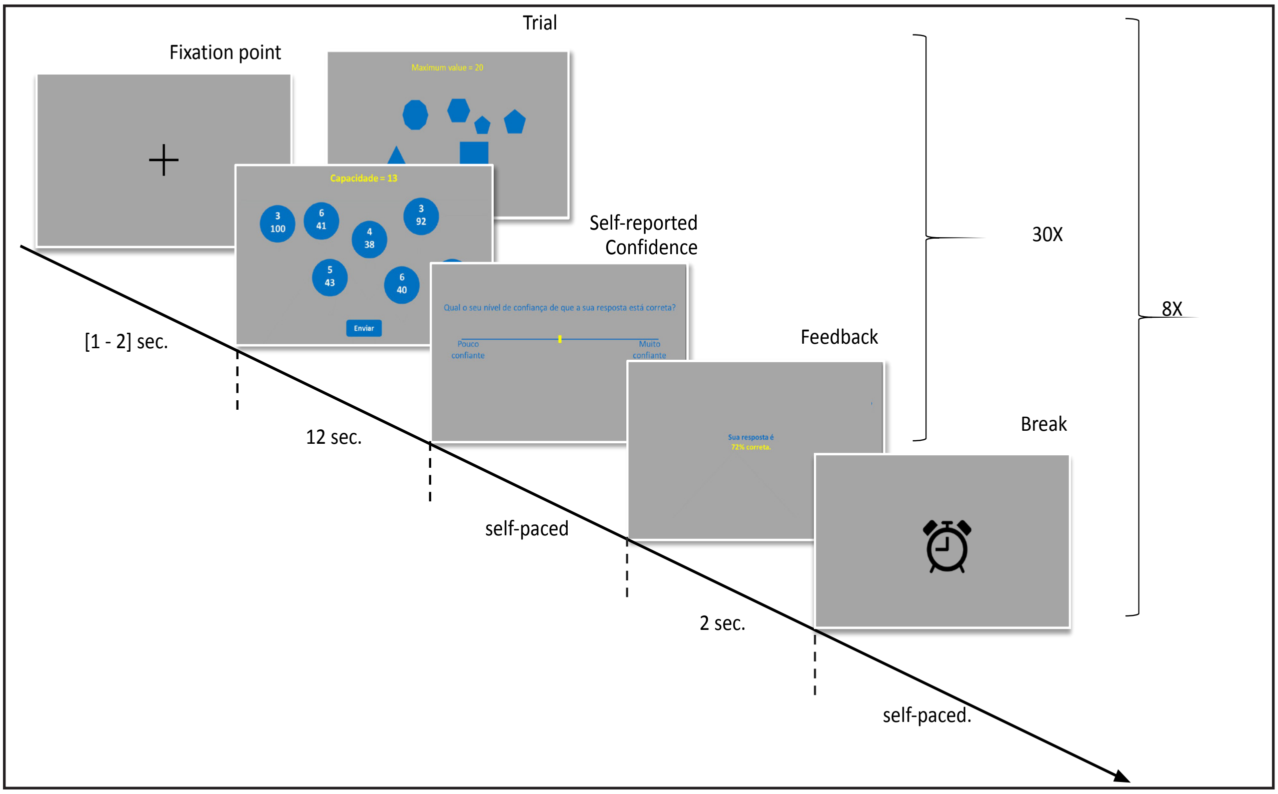



Note: In the numeric version, the items were circles of the same size, with the upper numbers representing the value and the lower numbers representing the weight. In the visual version, the items were polygons, with the number of sides representing the value and the size representing the weight. The figure shows equivalent items for both versions.
Source: Prepared by the authors

Participants were instructed to select, within a maximum of 12 seconds, a subset of items that maximized the total weight ($\sum_{i=1}^{n} w_i x_i$), without exceeding the imposed value constraint ($\sum_{i=1}^{n} v_i x_i \leq R$) .

A warning beep was played during the final 5 seconds of each trial to signal the imminent end of the time window, allowing participants to finalize their selections and submit a response. If no response was submitted within the 12 seconds, participants were informed that the attempt had been forfeited. Upon submission, participants were asked to indicate their confidence in the accuracy of their response. To maintain engagement and motivation, performance feedback was provided after each trial, following the confidence rating. The full structure of the task is illustrated in Figure 3.

Figure 3 – Task structure



Note: Before each trial, a fixation point was displayed for a randomized period between 1 and 2 seconds for subjects to focus their attention. Afterwards, an instance of the KP was displayed, for up to 12 seconds. Afterwards, confidence was self-reported using a slide bar (Not very confident – Very confident). The feedback was then displayed. 30 consecutive attempts were performed, followed by a break. Eight blocks were carried out.
Source: Prepared by the authors

Participants completed a total of 240 trials, organized into eight blocks. Within each block, all instances shared the same information presentation format, while task complexity factors varied both within and between blocks in a randomized manner. Additionally, all characteristics of the items comprising each instance were randomized to prevent potential confounds in task resolution. This included variation in item weights, values, restrictions, and the on-screen positioning of items. In accordance with the ethical guidelines of the laboratory in which the study was conducted, participants received monetary compensation for their involvement, consisting of a fixed amount of €8 and a variable bonus of up to €3, based on their average relative performance across trials.

Regarding the influence of participants' cognitive characteristics on task performance, Franco et al. (2021) employed tasks from the Cambridge Neuropsychological Test Automated Battery (CANTAB) to assess cognitive functioning and strategic processing in participants facing the same decision problem. Specifically, they used Reaction Time (RTI), Paired Associates Learning (PAL), and Spatial Working Memory (SWM) tests, which capture individual computational abilities such as mental arithmetic, working memory, episodic memory, strategy use, and psychomotor speed. The outcomes of these assessments were correlated with participants' performance in the experimental task on their study, but none of the correlations reached statistical significance. These results from Franco et al. (2021) suggest that the cognitive capacities evaluated did not constrain performance or account for its variability. Given these findings, and in view of the high implementation cost of the CANTAB battery as well as the already extended duration of the present study's experimental protocol - structured to ensure adequate statistical power - these cognitive assessments were not included in our data collection.

## 4 DATA ANALYSIS

The participants' ability to identify the optimal solution for each instance was assessed using the optimization performance variable. The overall average optimization performance was 52.38% (numerical version: 53.87%; visual version: 50.9%). A more flexible metric, relative performance, reflects how close the submitted solution was to the optimal one; the mean value for this variable was 0.89 in both versions.

Participants had a 12-second time limit per response, with an average response time (RT) of 8.44 seconds across subjects (numerical version: 8.91s; visual version: 7.97s). On average, participants reported being 77.13% confident in the correctness of their answers (numerical version: 77.26%; visual version: 77.0%).

In 3.15% of numerical trials and 1.72% of visual trials, participants failed to submit a response within the time limit. The value constraint was violated in 3.64% of numerical trials and 4.63% of visual trials. Following the procedure adopted by Franco et al. (2021), these non-compliant trials—where responses exceeded the maximum value or time—were excluded from hypothesis testing.

The effects of manipulating the number of items on each decision-making performance variable are presented in Table 3.

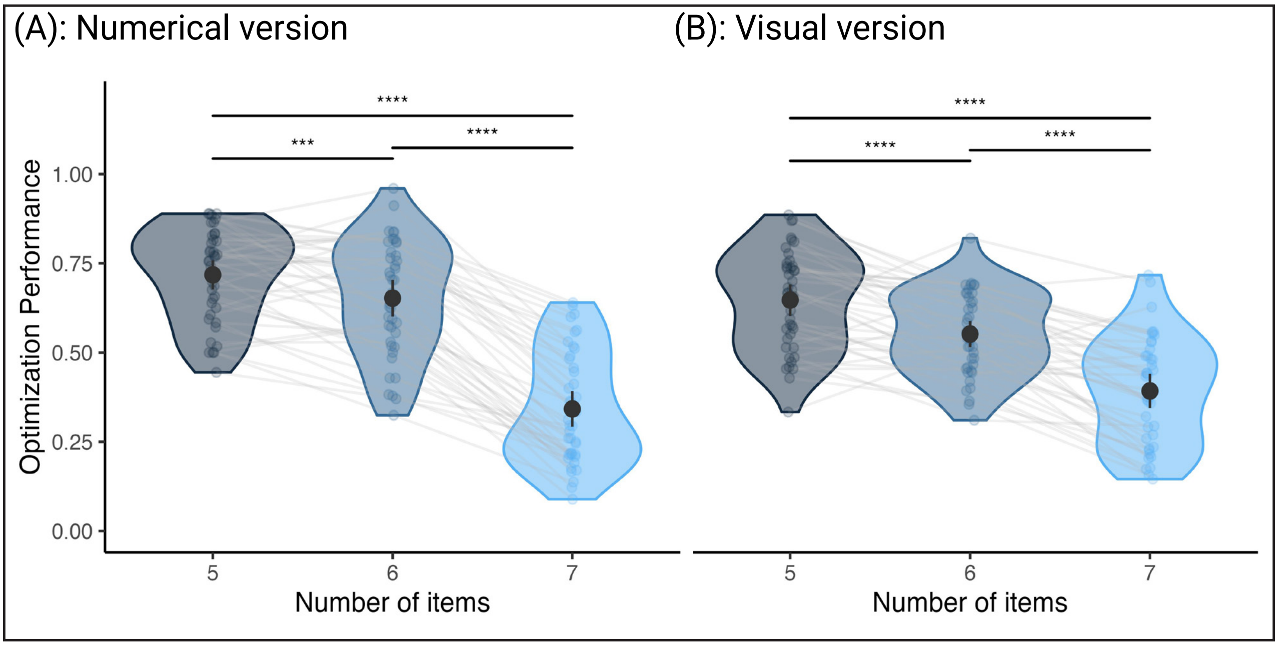
Table 3 – Statistical tests for the number of items

| Dependent variable | Statistical test | Presentation format | Statistics | *p-value* | Effect size/ Variance explained by the model |
|---|---|---|---|---|---|
| Optimization performance | Generalized Linear Mixed Model (GLMM) | *Numeric* | $X^2_{(2)} = 564.23$ | < 0.0001 | $R^2_{GLMM(c)} = 0.23$ |
| | | *Visual* | $X^2_{(2)} = 238.26$ | < 0.0001 | $R^2_{GLMM(c)} = 0.12$ |
| Relative performance | One-way repeated measures ANOVA | *Numeric* | F (2. 80) = 56.31 | < 0.0001 | $\eta_p^2 = 0.58$ |
| | Friedman's ANOVA | *Visual* | $X^2_{(2)} = 24.19$ | < 0.0001 | W = 0.29 |
| RT | One-way repeated measures ANOVA | *Numeric* | F (2. 80) = 214.32 | < 0.0001 | $\eta_p^2 = 0.84$ |
| | Friedman's ANOVA | *Visual* | $X^2_{(2)} = 70.58$ | < 0.0001 | W = 0.86 |
| Confidence | One-way repeated measures ANOVA with Grenhouse--Geisser correction | *Numeric* | F (2. 80) = 62.2 | < 0.0001 | $\eta_p^2 = 0.61$ |
| | | *Visual* | F (2. 80) = 45.67 | < 0.001 | $\eta_p^2 = 0.53$ |

Note: All tests conducted in R Studio. GLMM uses the "lme4" package, Wald Chi-square test uses the "car" package, one-way repeated measures ANOVA uses the "eZ" package, ηp2 effect size uses the "psychReport" package, Friedman's ANOVA uses the "Stats" package, Kendall effect size (W) uses the "rstatix" package.
Source: Prepared by the authors using research data

The GLMM results revealed a significant effect of the number of items on optimization performance. Each additional item was associated with a decrease in participants' ability to identify the optimal solution. In the numeric version, average optimization performance was 0.72 for instances with 5 items, 0.65 for 6 items, and 0.34 for 7 items. In the visual version, the respective values were 0.64, 0.56, and 0.39. These results are illustrated in Figure 4, which also displays the adjusted *p*-values from the post-hoc comparisons. All pairwise contrasts were statistically significant, thereby supporting Hypothesis 1 (H1).

Figure 4 – Impact from number of items on optimization performance

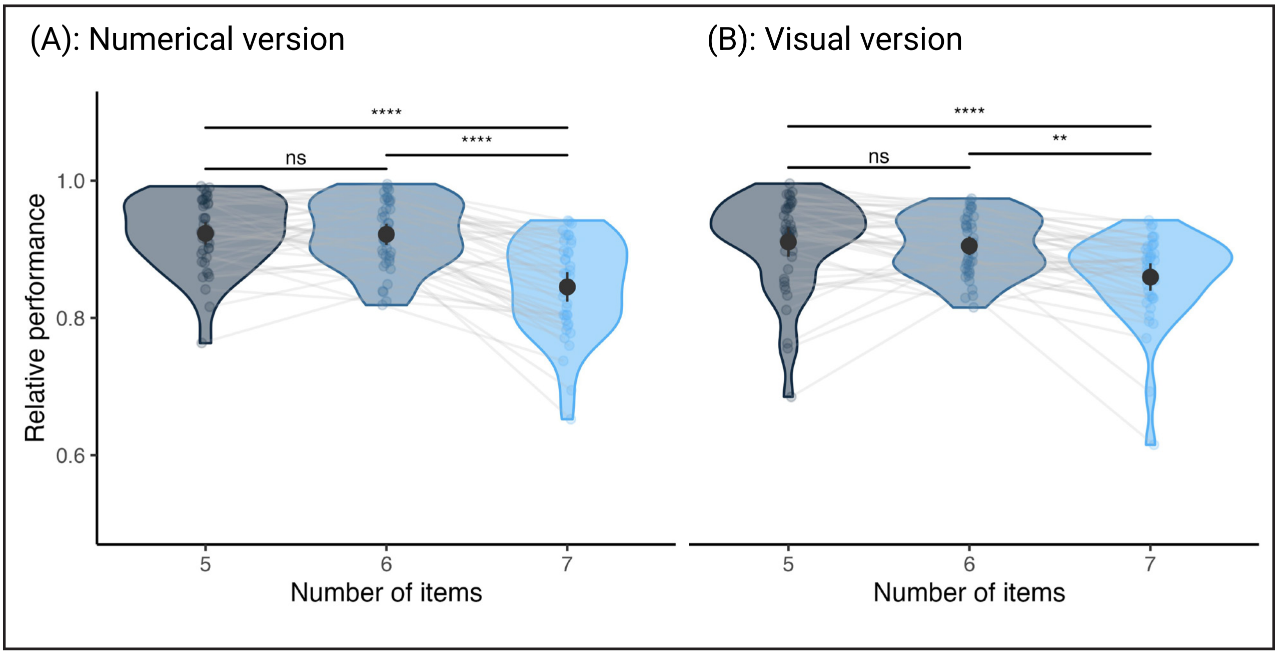



Note: Violin plots for optimization performance in numerical (A) and visual (B) versions. X-axis: Number of items in the instance (5, 6, 7). Colored dots represent the mean of each variable per subject; gray lines join the trials from each subject. Black dots represent the independent variable mean per level, along with black vertical bars representing the 95% confidence interval. Results of post-hoc tests presented by adjusted p-value ('***': p < 0.001; '****': p < 0.0001). All graphs generated in R Studio using the "ggplot2" package. Post-hoc test paired contrast with Bonferroni correction uses the "emmens" package.
Source: Prepared by the authors using research data

With respect to relative performance - a less stringent evaluation criterion - the results of the one-way repeated measures ANOVA (numerical condition) and Friedman's ANOVA (visual condition) revealed statistically significant differences across task sizes (Table 2; both $p < 0.0001$). However, a closer inspection of the descriptive statistics and post-hoc pairwise comparisons indicates that the observed performance decline is meaningfully concentrated in the most complex condition, involving 7 items.

In this case, participants' submitted responses deviated more substantially from the optimal solutions, providing partial empirical support for H2. As illustrated in Figure 5, mean relative performance remained stable between the 5- and 6-item conditions (numerical: 0.92; visual: 0.91), but dropped noticeably in the 7-item condition (numerical: 0.85; visual: 0.86).

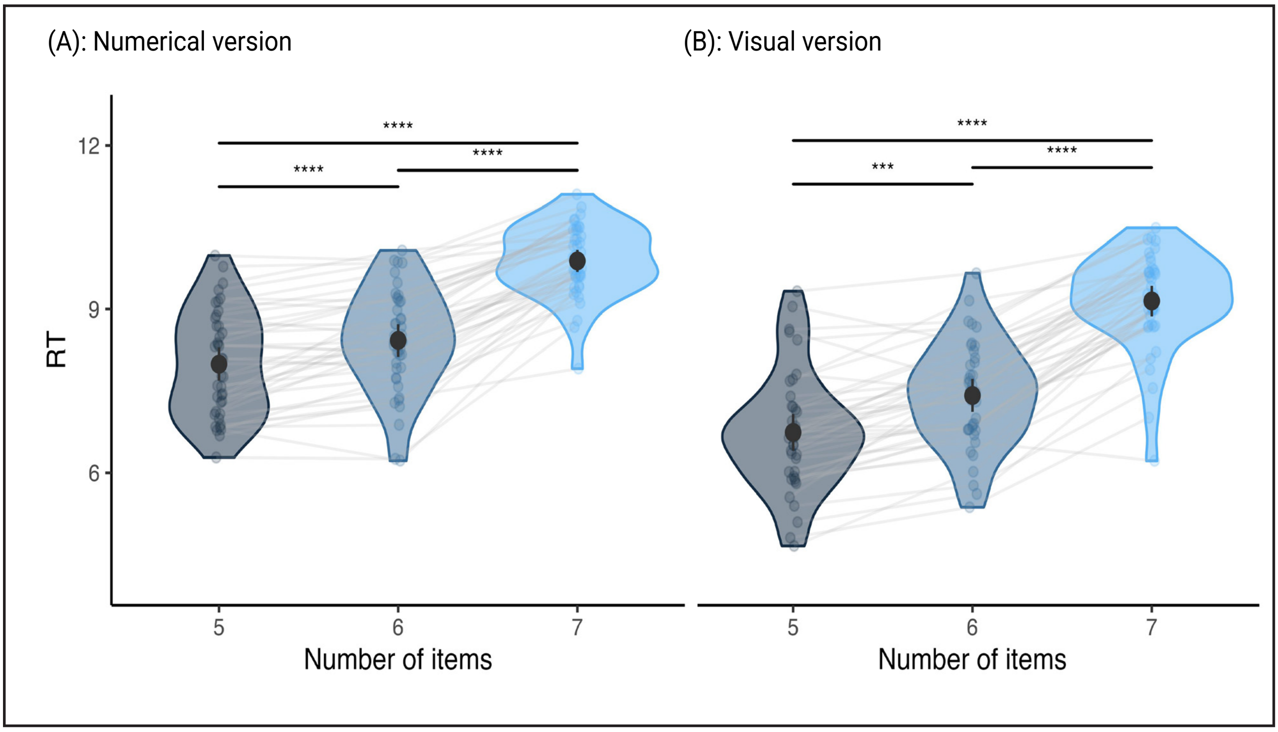
Figure 5 – Impact from number of items on relative performance




Note: Violin plots for relative performance in the numerical (A) and visual (B) versions. X-axis: Number of items in the instance (5, 6, 7). Colored dots represent the mean of each variable per subject; gray lines join the trials from each subject. Black dots represent the independent variable mean per level, along with black vertical bars representing the 95% confidence interval. Results of post-hoc tests presented by adjusted p-value ('ns': p > 0.05; '**': p < 0.01; '****': p < 0.0001). All graphs generated in R Studio using the "ggplot2" package. Paired post-hoc T-test with Bonferroni correction uses the "Stats" package and Siegel post-hoc test uses the "PMCMRplus" package.
Source: Prepared by the authors using research data

Response time (RT) varied systematically with the number of items per instance, as shown in Table 2. In the numerical condition, the mean RTs for instances with 5, 6, and 7 items were 7.96, 8.43, and 9.86 seconds, respectively. In the visual condition, the corresponding means were 6.75, 7.42, and 9.14 seconds. All pairwise comparisons were statistically significant, as illustrated in Figure 6, providing support for H3. It is also worth noting that the descriptive statistics, together with visual inspection of Figure 6, indicate that the increase in RT was greater between 6 and 7 items (numerical: +1.43s; visual: +1.72s) than between 5 and 6 items (numerical: +0.47s; visual: +0.67s).

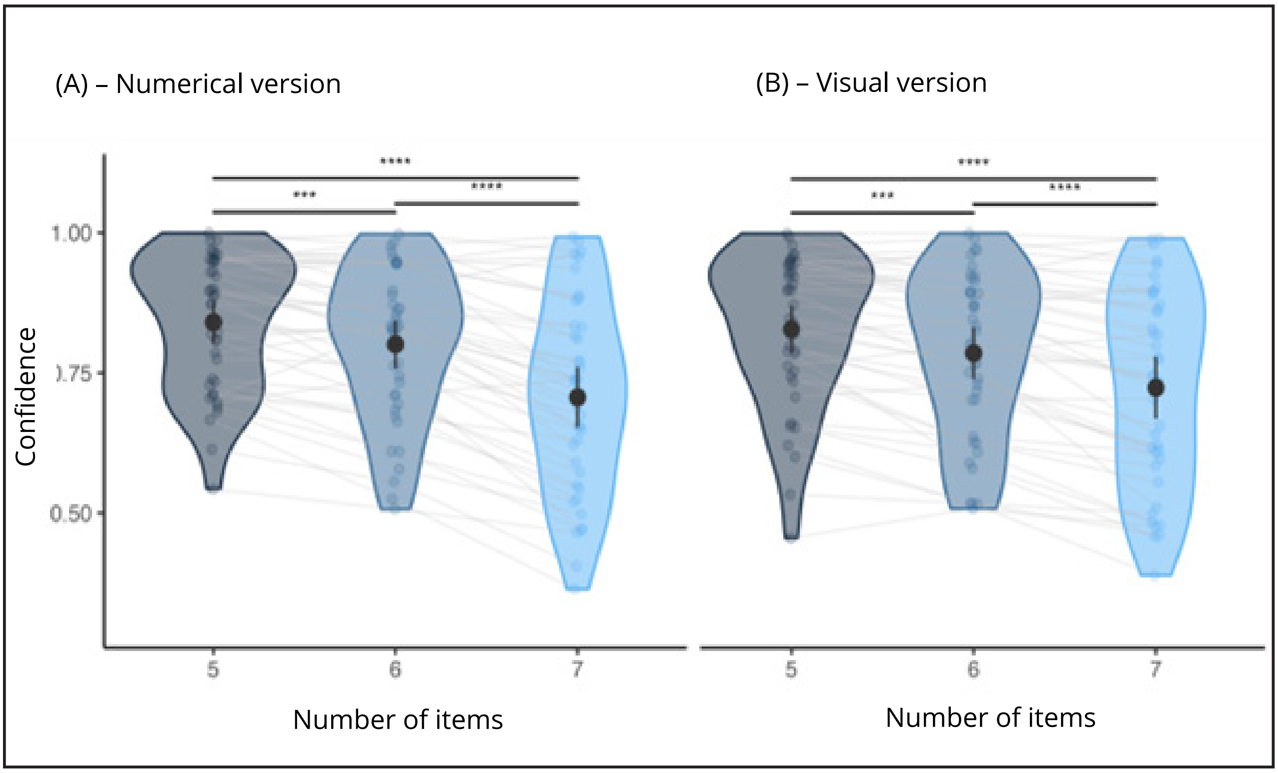
Figure 6 – Impact from the number of items on RT




*Note:* Violin plots for RT in numerical (A) and visual (B) versions. X-axis: Number of items in the instance (5, 6, 7). Colored dots represent the mean of each variable per subject; gray lines join the trials from each subject. Black dots represent the independent variable mean per level, along with black vertical bars representing the 95% confidence interval. Results of post-hoc tests presented by adjusted p-value ('****': p < 0.0001). All graphs generated in R Studio using the "ggplot2" package. Paired post-hoc T-test with Bonferroni correction uses the "Stats" package and Siegel post-hoc test uses the "PMCMRplus" package.
Source: Prepared by the authors using research data

As shown in Table 4, the one-way repeated measures ANOVA revealed a significant effect of the number of items on participants' decision confidence, measured immediately after the response submission and prior to any performance feedback. Confidence levels decreased as the number of items increased, a pattern clearly illustrated in Figure 7. Pairwise comparisons across all levels of the independent variable were statistically significant for both task versions, providing empirical support for H4. In the numerical condition, the mean confidence ratings were 0.84 for 5 items, 0.80 for 6 items, and 0.70 for 7 items. In the visual condition, the corresponding values were 0.83, 0.79, and 0.72.

Figure 7 – Impact from the number of items on confidence




*Note:* Violin plots for confidence in the numerical **(A)** and visual **(B)** versions. **X-axis**: Number of items in the instance (5, 6, 7). Colored dots represent the mean of each variable per subject; gray lines join the trials from each subject. Black dots represent the independent variable mean per level, along with black vertical bars representing the 95% confidence interval. Results of post-hoc tests presented by adjusted p-value ('***': $p < 0.001$; '****': $p < 0.0001$). All graphs generated in R Studio using the "ggplot2" package. Paired post-hoc T-test with Bonferroni correction uses the "Stats" package.
Source: Prepared by the authors using research data

Grounded in Computational Complexity Theory, the measures of input size and instance correlation capture distinct structural properties of the instances and reflect the computational difficulty associated with deriving a solution. These measures also enable an examination of how participants' responses relate to the intrinsic complexity of the task.

To assess the linear association between these complexity metrics and performance outcomes, repeated measures correlation was employed, as it appropriately adjusts Pearson and Spearman correlation coefficients for within-subject experimental designs (Table 4).

Table 3 – Input size linear correlation with dependent variables

| Dependent variable | Presentation format | | | |
|---|---|---|---|---|
| | ***numeric*** | | ***visual*** | |
| | ***p-value*** | $r_{rm}$ | ***p-value*** | $r_{rm}$ |
| Optimization performance | < 0.0001 | - 0.51 | < 0.0001 | - 0.32 |
| Relative performance | < 0.0001 | - 0.24 | < 0.0001 | - 0.17 |
| RT | < 0.0001 | + 0.72 | < 0.0001 | + 0.72 |
| Confidence | < 0.0001 | - 0.43 | < 0.0001 | - 0.36 |

Note: Tests conducted in R Studio. Repeated measure correlation uses the “rmcorr” package.
Source: Prepared by the authors using research data

Input size was defined as a joint function of two instance features: the number of items and the degree of value restriction. The results indicate that input size exerted a statistically significant effect on all dependent variables analyzed: it was associated with decreased optimization and relative performance, increased RT, and reduced decision confidence. These findings provide support for hypotheses H5, H6, H7, and H8.

With respect to optimization performance, the strength of the association with input size was moderate in the numerical condition and weak in the visual condition. For relative performance, the correlations were weak in the numerical condition and very weak in the visual condition. RT showed a strong positive correlation with input size in both task versions. Decision confidence exhibited a negative association, with moderate strength in the numerical version and weak strength in the visual version.

Regarding the instance correlation measure - intended to capture the extent to which item similarity impacts participants’ performance - significant effects were observed only for optimization performance and RT in the numerical condition, and for relative performance in the visual condition (see Table 5). These results provide partial support for H9 and H11 (numerical version only), and for H10 (visual version only). H12 is not supported, as the corresponding p-values exceeded the 0.05 threshold.

It is important to note, however, that even in cases where statistical significance was found, the observed correlations were very weak. This suggests that while instance correlation has some measurable influence on these dependent variables, additional factors likely contribute to the relationships and warrant further investigation.

Table 5 – Instance correlation linear correlation with dependent variables

| Dependent variable | Presentation format | | | |
|---|---|---|---|---|
| | *numeric* | | *visual* | |
| | *p-value* | $r_{rm}$ | *p-value* | $r_{rm}$ |
| Optimization performance | < 0.001 | - 0.05 | > 0.05 | - 0.02 |
| Relative performance | > 0.05 | - 0.02 | < 0.0001 | - 0.06 |
| RT | < 0.05 | + 0.03 | > 0.05 | - 0.06 |
| Confidence | > 0.05 | 0.00 | > 0.05 | 0.00 |

Note: Tests conducted in R Studio. Repeated measure correlation uses the “rmcorr” package.
Source: Prepared by the authors using research data

The results of the hypothesis tests presented in this section are discussed in the next section in accordance with the theoretical and empirical literature.

## 5 DISCUSSION

One of the most fundamental intrinsic properties of a task that can increase the complexity of decision-making and problem-solving is the number of items or elements involved. This has been well established in the task complexity literature (Campbell, 1988; Liu & Li, 2012; Mackinnon & Wearing, 1980; Simon, 1962; Wood, 1986), as exceeding an individual’s cognitive processing limits results in additional elements becoming a cognitive burden (Sweller et al., 2019).

From the perspective of Computational Complexity Theory, an increase in the number of items in a combinatorial optimization problem corresponds to an exponential growth in the number of candidate solutions (Bossaerts & Murawski, 2017; Franco et al., 2021).

The results of the present study demonstrate a robust effect of item quantity on task performance. In both task versions, increasing the number of items led to a decline in the ability to identify the optimal solution, as measured by optimization performance, and also impaired participants' ability to provide an approximate solution, as indicated by their relative performance. Additionally, higher numbers of items were associated with increased response times and decreased self-reported confidence.

Consistent with these findings, Murawski and Bossaerts (2016), using a numerical version of a similar task with 10 and 12 items, also observed a negative impact on both optimization and relative performance, indicating that tasks with more items are inherently more difficult for human solvers. The current evidence supports their conclusion that human difficulty correlates with mathematically defined instance complexity, such as that described by Turing machine models. Similar effects have been previously observed in visual optimization tasks like the Traveling Salesman Problem (Hidalgo-Herrero et al., 2013). The present study extends this evidence by demonstrating the same pattern in a visual version of the Knapsack Problem (KP).

In terms of time, the findings confirm that increasing the number of elements in a problem instance leads to a higher number of operations required for its resolution, characterizing an increase in time complexity (Bossaerts & Murawski, 2017), given that the number of items directly determines the number of feasible solutions.

The current results align with those of Pizlo and Zhen Li (2005), who found that solving the combinatorial n-puzzle problem[l] with 5, 8, 15, and 35 pieces required progressively more time as the number of pieces increased.

With respect to relative performance, no statistically significant difference was observed between the 5- and 6-item conditions in either version of the task. However, for the RT, the increase from 5 to 6 items yielded a significant effect. These findings suggest that working memory could accommodate the additional item without compromising solution quality, though at the cost of increased time investment. This indicates a compensatory strategy may have been employed, whereby participants devoted more time to maintain suboptimal performance levels.

In contrast, all statistical comparisons between the 6- and 7-item instances revealed significant differences, including reductions in relative performance and

[l] Classic puzzle game where you have n+1 spaces and you must move disordered pieces in to order them.

a marked increase in RT. These results imply that the compensatory mechanism observed for the 6-item condition became unsustainable. The additional cognitive load associated with 7 items could no longer be offset by increased time, indicating the emergence of a performance threshold or phase transition, beyond which cognitive resources were insufficient to maintain task effectiveness.

These findings are consistent with the seminal work of Miller (1956), as further discussed by Simon (1990), which provided empirical support for Sweller's (1988) cognitive load theory. Miller (1956) famously proposed that working memory has a limited capacity of approximately 7 ± 2 informational units, with the variability attributed to the complexity of the material being processed.

In this context, the computational complexity measures used in the task - namely input size and instance correlation - are grounded in the algorithmic principles employed by standard Knapsack Problem (KP) solvers. The present study investigated whether variations in these computational features would correspond with behavioral performance, thereby suggesting that participants may have adopted strategies analogous to those used in algorithmic approaches.

As outlined by Murawski and Bossaerts (2016), input size reflects the demands of dynamic programming algorithms, which operate on sets of heuristic rules that, while effective, are computationally intensive and require working memory resources beyond typical human capacity. In contrast, instance correlation captures the structure exploited by greedy algorithms, which prioritize items according to a ratio - typically value-to-weight or, in this study, weight-to-value given the constraint applied to item value. Identifying such correlations between algorithmic complexity and human performance provides insights into the cognitive strategies likely adopted by participants under varying task demands.

The linear correlation analyses indicate that *input size* - defined as a function of the number of items and the imposed value constraint - was more strongly associated with the performance variables in decision-making. Although this reflects a computationally demanding strategy, participants generally began their selection process with items of lower weight, implicitly aiming to maximize value through such choices.

For more complex instances - i.e., those with larger input sizes - this approach was associated with diminished performance and increased response times. This pattern was consistent across both task versions. However, the relationship is only partially explanatory; the highest coefficient of determination ($R^2 = 0.5184$) suggests that input size accounts for approximately 51.84% of the variation in RT, indicating that additional factors are likely influencing this relationship.

By contrast, instance correlation showed a statistically significant but consistently weak association, based on effect size criteria established by Cohen (1988). These findings suggest that participants, in general, did not apply a strategy based on ranking items by value-to-weight ratio. This was expected in the visual version, where discrete weight values were not accessible to participants. However, even in the numerical version, where such a strategy was feasible, it was not employed.

We propose that the time constraint imposed - although not constituting formal time pressure as in previous validations of the task - led participants to perceive insufficient time to carry out the necessary calculations. Participants likely recognized the limitations of their working memory in rapidly storing and manipulating numerical information.

Computing the ratio of value (ranging from 3 to 10) to weight (ranging from 20 to 99) invariably yields decimal values, making the process cognitively demanding. These results support the claims of Murawski and Bossaerts (2016), who argue that algorithm-derived strategies involving arithmetic operations may exceed human cognitive capacity, particularly when tasks require division of complex numerical sequences.

Nevertheless, Murawski and Bossaerts (2016) reported a significant negative association between instance correlation and optimization performance, alongside an absence of association with input size. This partially contrasts with the current findings, which identified a significant negative correlation with input size, but only a weak and, in the case of the visual version, nonsignificant association with instance correlation. These discrepancies may reflect differences in task design between the current study and that of Murawski and Bossaerts (2016).

## 6 FINAL CONSIDERATIONS

This study aimed to examine how factors contributing to the complexity of an optimization problem affect human decision-making performance. To this end, a task based on the Knapsack Problem (KP) - a well-established combinatorial optimization problem with strong ecological validity - was employed. The KP was selected for its relevance to real-world decision contexts, its suitability for experimental manipulation in both laboratory and field settings, and its capacity to isolate multiple complexity-related factors, thus addressing gaps in the existing literature.

Specifically, three factors known to influence problem complexity were investigated: number of items, input size, and instance correlation. Their impact was assessed across four outcome variables: the ability to identify the optimal solution (optimization performance), the ability to generate approximate solutions (relative performance), response time (RT), and decision confidence (via self-report).

The manipulation of the number of items yielded robust effects across both the numerical and visual versions of the task. Instances with a greater number of items were associated with decreased optimization and relative performance, increased RT, and lower confidence ratings. Notably, a phase transition in relative performance was identified at the 7-item level. While participants were able to compensate for the increased cognitive load from 5 to 6 items by extending RT, this compensatory strategy failed between 6 and 7 items.

This suggests a cognitive threshold beyond which additional processing time is insufficient to maintain performance, leading to a marked decline in solution quality. Furthermore, the observed decline in confidence as a function of item number provides novel evidence regarding metacognitive aspects of decision-making in complex tasks - a variable not previously examined in this context.

These results support hypothesis H1, H2, and H4, which predicted negative effects of item number on optimization performance, RT, and confidence, respectively. Hypothesis H3, regarding relative performance, was partially supported.

With respect to the measures derived from algorithmic difficulty - specifically, input size and instance correlation - the results revealed a significant linear correlation between input size and all assessed indicators of decision-making performance: optimization performance, relative performance, RT, and confidence. These findings lend support to hypotheses H5, H6, H7, and H8, respectively.

Notably, the strength of association was highest for RT, indicating that input size had a particularly strong effect on the time participants required to submit a response. This suggests that, overall, participants may have adopted a heuristic strategy centered on selecting items of lower weight first - likely aiming to maximize total value as a secondary consequence of this selection pattern. Among all performance metrics, the relationship between input size and RT was the most pronounced, with input size accounting for 51.84% of the variance in response time.

In contrast, instance correlation showed limited influence. Partial support was found only for hypotheses H9 and H11 - those pertaining to optimization performance and RT - in the numerical version of the task. For the visual version, partial support emerged for H10, related to relative performance. However, in neither version was there sufficient evidence to support the hypothesis (H12) regarding the effect of instance correlation on decision confidence.

Furthermore, even where significance was achieved, the strength of association was consistently weak, suggesting that participants did not adopt the computationally intensive strategy of dividing value by weight to derive an index for item ranking. This finding is consistent with cognitive load limitations and aligns with prior research suggesting that arithmetic-based strategies involving division are not typically employed under bounded time and memory constraints.

Taken together, these findings indicate that, for the optimization problem employed in this study, the number of items and their interaction with internal problem characteristics - specifically the imposed value constraint, operationalized via the input size metric - were the most impactful contributors to perceived task complexity.

This reinforces the established importance of the number of elements or items in influencing task difficulty, a conclusion supported by multiple bodies of literature, including decision-making (Payne, 1976; Payne et al., 1988, 1991, 1993; Tversky & Kahneman, 1981), task complexity (Campbell, 1988; Liu & Li, 2012; Mackinnon & Wearing, 1980; Simon, 1962; Wood, 1986), cognitive load theory (Mangion, 2017; Sweller, 1988, 2006), and computational complexity (Carruthers et al., 2012; Franco et al., 2021; Hidalgo-Herrero et al., 2013; MacGregor & Chu, 2011; Murawski & Bossaerts, 2016; Pizlo & Zhen Li, 2005).

This conclusion is further substantiated by the robust effects of item number on performance when examined independently, as well as by the disproportionately large influence this variable exerts within the input size function (defined as ) where the number of items contributes more substantially than the logarithmically scaled value constraint.

Theoretically, this research establishes a conceptual bridge between task complexity literature - focused on attributes such as quantity, diversity, and redundancy - and measures derived from Computational Complexity Theory, which capture the impact of specific instance characteristics and reflect the strategies used in problem-solving.

While these domains have long overlapped, they had not previously been jointly evaluated in a manner that enables a more comprehensive understanding of their combined effects on human decision-making. A further theoretical contribution lies in integrating computational complexity measures into decision-making research. By grounding the complexity classification in a rigorous mathematical framework, this approach broadens the theoretical base applicable to managerial and behavioral investigations.

Specifically, the input size measure proved suitable for the problem studied and offers a means to quantify both the intrinsic complexity of the task and the computational demands imposed on the decision-maker. This presents a promising avenue for future research aiming to dissect complexity in tasks and problem-solving contexts.

The findings of this study offer important insights for professionals and managers involved in contexts that require solving complex decision problems, particularly those involving resource allocation, prioritization, and strategic planning. By demonstrating that the number of elements and their interaction with task constraints (as captured by input size) significantly affect performance, RT, and confidence, this research highlights specific conditions under which decision quality is likely to deteriorate.

Managers should be aware that increasing the number of elements in a task—even by a single unit—can trigger a disproportionate increase in cognitive load and decision time, potentially compromising the accuracy of decisions. In operational and strategic settings such as budgeting, portfolio selection, or production scheduling, decision support systems should be designed to present manageable subsets of information at each stage, rather than overwhelming the user with all variables at once.

The significant effect of input size on response time and performance suggests that professionals often attempt to compensate for complexity by investing more time in the task—up to a point. However, when complexity exceeds individual cognitive limits, performance drops regardless of time invested. To prevent this, managers should implement structural task simplifications, such as reducing the number of simultaneous options, using pre-filtered alternatives, or automating part of the evaluation process in complex decision environments.

In systems where humans interact with algorithmic recommendations—such as financial platforms, supply chain systems, or AI-enhanced dashboards—the complexity of the instance should inform the way information is presented. For high-complexity problems, systems could dynamically adjust the granularity of information, highlight key trade-offs, or visually represent constraint violations to support more effective human judgment.

Despite the methodological rigor adopted in the present study, several limitations should be acknowledged. First, the sample consisted predominantly of young adults, likely university students, which may constrain the generalizability of the findings to broader populations. Age and educational background can influence cognitive performance, particularly in tasks involving digital interfaces and time-constrained problem solving.

Second, the experimental sessions lasted between 40 minutes and one hour, which, given the cognitively demanding nature of the task, could have introduced participant fatigue, especially in the later stages of the procedure. Moreover, an examination of performance trends throughout the task revealed no evidence of deterioration as the task progressed. However, this concern was not supported by post-experiment reports. At the end of the study, participants were asked whether they experienced fatigue, and none reported feeling tired. On the contrary, most expressed enthusiasm, describing the task as game-like and reporting that the immediate performance feedback was motivating, encouraging them to improve their responses to increase their final compensation. Based on these self-reports, there is no indication that fatigue negatively impacted performance.

Additionally, individual differences in working and episodic memory, strategic planning, processing speed, psychomotor function, and mathematical ability were not directly assessed. Although Franco et al. (2021) found no significant correlations between such cognitive capacities and performance on a similar task, it is possible that studies conducted with more diverse samples - particularly across different age groups or cultural backgrounds - may reveal meaningful variability. Future research should consider incorporating cognitive assessments and recruiting heterogeneous samples in terms of age, education, and professional experience to enhance the external validity and explanatory scope of the findings.

Building on the limitations identified in the present study, future research is encouraged to address several opportunities for advancement. While the controlled laboratory setting enabled precise manipulation of task complexity, it may reduce ecological validity; subsequent studies should test whether the observed effects hold in real-world contexts such as organizational decision-making, where uncertainty, competing goals, and time pressure are more pronounced.

The sample, composed mainly of young and educated participants, suggests the need to replicate the findings in more diverse populations, including professional decisionmakers. Additionally, while the study demonstrated how complexity impacts performance, it did not directly investigate the cognitive strategies employed by participants.

Future research should incorporate qualitative or quantitative assessments - such as post-experimental questionnaires or analyses of item selection patterns - to better understand how individuals navigate complex decision tasks. Such analyses could clarify the relationship between strategy use, performance, and response time, and reveal whether these strategies differ across presentation formats (e.g., numerical vs. visual), thus enriching our understanding of decision behavior under varying cognitive demands. Finally, future experiments should explore a broader set of instance-level features, such as variations in constraint tightness and item value distributions, to deepen understanding of how specific problem structures influence cognitive effort and decision quality.

## REFERENCES

Adomavicius, G., Curley, S. P., Gupta, A., & Sanyal, P. (2020). How decision complexity affects outcomes in combinatorial auctions. *Production and Operations Management*, 29(11), 2579-2600. https://doi.org/10.1111/poms.13249

Aitchison, L., Bang, D., Bahrami, B., & Latham, P. E. (2015). Doubly Bayesian analysis of confidence in perceptual decision-making. *PLoS computational biology*, *11*(10), e1004519.

Bossaerts, P., & Murawski, C. (2017). Computational Complexity and Human Decision-Making. *Trends in Cognitive Sciences*, *21*(12), 917–929. https://doi.org/10.1016/j.tics.2017.09.005

Brus, J., Aebersold, H., Grueschow, M., & Polania, R. (2021). Sources of confidence in value-based choice. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-27618-5

Burns, N. R., Lee, M. D., & Vickers, D. (2006). Are Individual Differences in Performance on Perceptual and Cognitive Optimization Problems Determined by General Intelligence? *The Journal of Problem Solving*, *1*(1), 5–19. https://doi.org/10.7771/1932-6246.1003

Campbell, D. J. (1988). Task Complexity: A Review and Analysis. *Academy of Management Review*, *13*(1), 40–52. https://doi.org/10.5465/amr.1988.4306775

Carruthers, S. E. (2015). *The Role of the Goal in Problem Solving Hard Computational Problems: Do People Really Optimize?*

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.

Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, *8* (jul), 1–11. https://doi.org/10.3389/fpsyg.2017.01153

Double, K. S., & Birney, D. P. (2018). Reactivity to confidence ratings in older individuals performing the latin square task. *Metacognition and Learning*, *13*(3), 309–326. https://doi.org/10.1007/s11409-018-9186-5

Erdfelder, E., FAul, F., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Franco, J. P., Doroc, K., Yadav, N., Bossaerts, P., & Murawski, C. (2022). *Task-independent metrics of computational hardness predict human cognitive performance*. https://doi.org/10.1101/2021.04.25.441300

Franco, J. P., Yadav, N., Bossaerts, P., & Murawski, C. (2021). Generic properties of a computational task predict human effort and performance. *Journal of Mathematical Psychology*, *104*. https://doi.org/10.1016/j.jmp.2021.102592

Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology*, *83*(3), 269–275. https://doi.org/10.1016/j.ijpsycho.2011.09.023

Gavas, R. D., Tripathy, S. R., Chatterjee, D., & Sinha, A. (2018). Cognitive load and metacognitive confidence extraction from pupillary response. *Cognitive Systems Research*, *52*, 325–334. https://doi.org/10.1016/j.cogsys.2018.07.021

Hidalgo-herrero, M., Rabanal, P., Rodríguez, I., & Rubio, F. (2013). *Comparing Problem Solving Strategies for NP-hard*. *124*, 1–25. https://doi.org/10.3233/FI-2013-822

Klepsch, M., & Seufert, T. (2020). Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. Em *Instructional Science* (Vol. 48, Número 1). Springer Netherlands. https://doi.org/10.1007/s11251-020-09502-9

Lempert, K. M., Chen, Y. L., & Fleming, S. M. (2015). Relating pupil dilation and metacognitive confidence during auditory decision-making. *PLoS ONE*, *10*(5). https://doi.org/10.1371/journal.pone.0126588

Liu, P., & Li, Z. (2012). Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics*, *42*(6), 553–568. https://doi.org/10.1016/j.ergon.2012.09.001

Mackinnon, A. J., & Wearing, A. J. (1980). Complexity and decision making. *Behavioral Science*, *25*(4), 285–296. https://doi.org/10.1002/bs.3830250405

Mangion, K. L. H. (2017). *Human cognitive performance: a neurophysiological assessment of the impact that reverse assessment priming has on mental workload, performance and cognitive efficiency during transient information processing*. University of Southern Queensland.

Meloso, D., Copic, J., & Bossaerts, P. (2009). Promoting intellectual discovery: Patents versus markets. *Science*, *323*(5919), 1335–1339. https://doi.org/10.1126/science.1158624

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev*, *63*(2), 81–97. https://doi.org/10.1177/001088049003100202

Murawski, C., & Bossaerts, P. (2016). How Humans Solve Complex Problems: The Case of the Knapsack Problem. *Scientific Reports*, *6*(September), 1–10. https://doi.org/10.1038/srep34851

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, *16*(2), 366–387. https://doi.org/10.1016/0030-5073(76)90022-2

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534–552. https://doi.org/10.1037/0278-7393.14.3.534

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1991). Consumer decision making. Em *Handbook of consumer behaviour* (p. 50–84). Prentice-Hall.

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

Pizlo, Z., & Zhen Li. (2005). Solving combinatorial problems: The 15-puzzle. *Memory & Cognition*, *33*(6), 1069–1084.

Ross, N. D. F., Johns, M. B., Keedwell, E. C., & Savic, D. A. (2019). Human-evolutionary problem solving through gamification of a bin-packing problem. *GECCO 2019 Companion - Proceedings of the 2019 Genetic and Evolutionary Computation Conference Companion*, 1465–1473. https://doi.org/10.1145/3319619.3326871

Simon, H. A. (1962). The Architecture of Complexity Proceedings of the American Philosophical Society Vol. 106.

Simon, H. A. (1990). Invariants of human behavior. *Annual review of psychology*, *41*(1), 1-20.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Wei, Z., & Wang, X. J. (2015). Confidence estimation as a stochastic process in a neurodynamical system of decision making. *Journal of neurophysiology*, *114*(1), 99-113. https://doi.org/10.1152/jn.00793.2014

Wood, R. E. (1986). Task Complexity: Definition of the Construct. *Organizational behavior and human decision processes*, *37*, 60–82.

Zagermann, J., Pfeil, U., & Reiterer, H. (2016). Measuring cognitive load using eye tracking technology in visual computing. *ACM International Conference Proceeding Series*, *24-October*, 78–85. https://doi.org/10.1145/2993901.2993908

## Authors

### 1 – Carolina Schneider Bender

Institution: Universidade Federal de Santa Maria – Santa Maria, Rio Grande do Sul, Brazil

Doctor of Administration from Federal University of Santa Maria (UFSM)

Orcid: https://orcid.org/0000-0001-9234-7447

E-mail: carolisbender@gmail.com

### 2 – Mauri Leodir Löbler

Institution: Universidade Federal de Santa Maria – Santa Maria, Rio Grande do Sul, Brazil

Doctor of Administration from Federal University of Rio Grande (UFRGS)

Orcid: https://orcid.org/0000-0002-0244-6351

E-mail: mllobler@gmail.com

### 3 – Eliete dos Reis Lehnhart

Institution: Universidade Federal de Santa Maria – Santa Maria, Rio Grande do Sul, Brazil

Doctor of Administration from Federal University of Santa Maria (UFSM)

Orcid: https://orcid.org/0000-0001-6676-8547

E-mail: eliete.lehnhart@ufsm.br

## Contribution of authors

| Contribution | [Author 1] | [Author 2] | [Author 3] |
|---|---|---|---|
| 1. Definition of research problem | √ | √ | √ |
| 2. Development of hypotheses or research questions (empirical studies) | √ | √ | √ |
| 3. Development of theoretical propositions (theoretical work) | √ | √ | √ |
| 4. Theoretical foundation / Literature review | √ | √ | |
| 5. Definition of methodological procedures | √ | √ | √ |
| 6. Data collection | √ | | |
| 7. Statistical analysis | √ | | |
| 8. Analysis and interpretation of data | √ | √ | √ |
| 9. Critical revision of the manuscript | √ | √ | √ |
| 10. Manuscript writing | √ | √ | √ |
| 11. Other (please specify) | | | |

***Conflict of Interest***

*The authors have stated that there is no conflict of interest.*



***Plagiarism Check***

*The ReA/UFSM maintains the practice of submitting all documents approved for publication to the plagiarism check, using specific tools, e.g.: Turnitin.*

***Edited by***

*Jordana Marques Kneipp*