Kivikangas, J. Matias; Vilkkumaa, Eeva; Blank, Julian; Harjunen, Ville; Malo, Pekka; Deb, Kalyanmoy; Ravaja, Niklas J.; Wallenius, Jyrki

**Effects of many conflicting objectives on decision-makers' cognitive burden and decision consistency**

*Published in:*
European Journal of Operational Research

*DOI:*
10.1016/j.ejor.2024.10.039

Published: 01/04/2025

*Document Version*
Publisher's PDF, also known as Version of record

European Journal of Operational Research 322 (2025) 182–197

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/eor

Decision Support

# Effects of many conflicting objectives on decision-makers' cognitive burden and decision consistency

J. Matias Kivikangas [a,b], Eeva Vilkkumaa [b,*], Julian Blank [c], Ville Harjunen [a], Pekka Malo [b], Kalyanmoy Deb [c], Niklas J. Ravaja [a], Jyrki Wallenius [b]

[a] *Department of Psychology and Logopedics, University of Helsinki, Finland*
[b] *Department of Information and Service Management, Aalto University School of Business, Finland*
[c] *Department of Computer Science and Engineering, Michigan State University, USA*

ARTICLE INFO

*Keywords:*
Multiple criteria analysis
Multiple Criteria Decision Making
Evolutionary Multi-Objective Optimization
Cognitive burden
Psychophysiological measurements

ABSTRACT

Practical planning and decision-making problems are often better and more accurately formulated with multiple conflicting objectives rather than a single objective. This study investigates a situation relevant for Multiple Criteria Decision Making (MCDM) as well as Evolutionary Multi-objective Optimization (EMO), where the decision-maker needs to make a series of choices between nondominated options characterized by multiple objectives. The cognitive capacity of humans is limited, which leads to *cognitive burden* that influences human decision-makers' decisions. We measure how the varying number of objectives influences cognitive burden in a laboratory study, and the impacts that this burden has on the decision-makers' behavior and the consistency of their decisions. We use psychophysiological, behavioral, and self-report methods. Our results suggest that a higher number of objectives (i) increases cognitive burden significantly, (ii) leads to adopting strategies in which only a limited number of objectives is considered, and (iii) decreases decision consistency.

## 1. Introduction

Many design, planning, and management problems are represented and solved as multiobjective decision-making or optimization problems (Deb, 2001; Korhonen & Wallenius, 2020; Miettinen, 1999; Wallenius et al., 2008). In such problems, the concept of an optimum (from single-objective optimization) is generalized to the set of Pareto-optimal or non-dominated solutions, which cannot be improved with respect to any objective without impairing performance on some other objective(s). Over half a century, numerous effective multiobjective decision-making and optimization algorithms have been developed to support decision-making in such problems. Much of the early work, which is referred to with the abbreviation MCDM (Multiple Criteria Decision Making[1]), is described in Köksalan, Wallenius, and Zionts (2011). The algorithms extend ideas presented in Decision Analysis or mathematical programming. Although the strategies vary, the purpose of such algorithms is to help decision-makers identify their most preferred solution among the set of Pareto-optimal solutions. The algorithms can be categorized as a priori, a posteriori, or interactive (progressive), depending on when preference elicitation takes place.

About 20–30 years ago, a related field, focusing on optimization under multiple objectives, called Evolutionary Multiobjective Optimization (EMO) emerged (Coello, 2007; Deb, Pratap, Agarwal, & Meyarivan, 2002; Fonseca & Fleming, 1993; Srinivas & Deb, 1994; Zitzler, 1999). The purpose of EMO algorithms is to identify a well-converged and well-diversified set of nearly Pareto-optimal solutions. The EMO approaches usually follow a population-based optimization approach (Deb, 2001), where the "fitness" of the population (of solutions) is improved from one generation to the next by mimicking ideas from nature. Once a representative set of near Pareto-optimal solutions is found, the task remains to identify the most preferred solution. At the beginning of EMO's growth, the emphasis was placed on solving two- and three-objective problems. However, for the past decade or so, much interest has been devoted to handling an increasing number of objectives, resulting in a new sub-field: Evolutionary Many-objective Optimization (EMaO) (Deb & Jain, 2013; Zhang & Li, 2007). Each new algorithm or test problem suite is expected to have considered many objectives (informally, 4–15 objectives) for the study to be of value.

From the perspective of human decision-makers, the difficulty lies in choosing one out of many Pareto-optimal solutions. Several preference

* Correspondence to: Aalto University School of Business, Department of Information and Service Management, P.O. Box 21210, 00076, Aalto, Finland.
*E-mail address:* eeva.vilkkumaa@aalto.fi (E. Vilkkumaa).

[1] The term "criterion" (as well as "attribute") is typically used to refer to a tangible element through which the attainment of an objective is measured, but the terms "objective", "criterion", and "attribute" are also sometimes used interchangeably.

https://doi.org/10.1016/j.ejor.2024.10.039
Received 26 September 2023; Accepted 28 October 2024
Available online 6 November 2024

elicitation techniques (including methods of interactive multiobjective optimization; see Afsar et al., 2022; Miettinen, Ruiz, & Wierzbicki, 2008) have been suggested to alleviate the cognitive burden associated with this task, but even these methods require that the decision-maker is able to compare multiobjective decision alternatives, which may be challenging when the number of objectives is large. This decision-making difficulty, involving human decision-makers, is largely ignored in the EMO and EMaO studies, but also to some extent in the MCDM studies. For instance, real-life applications of MCDM/EMO methods to environmental and engineering problems routinely consider 8–11 objectives (Linkov, Moberg, Trump, Yatsalo, & Keisler, 2020). Particularly, EMO/EMaO researchers have developed various objective-wise scalable problems to test the efficacy of their optimization algorithms (Deb et al., 2002; Huband, Barone, While, & Hingston, 2005; Zhang et al., 2008). Most recent academic studies include 10 or more objectives (test and engineering design problems) to demonstrate the "working" of proposed EMaO algorithms and also to meet increasing demand for "quality works" from researchers. It is rather commonly assumed that the more objectives the decision-maker must consider, the more difficult decision-making becomes, but the details have not been demonstrated clearly. In the worst case, human cognitive limitations may severely restrict the practical usability of these algorithms.

The impact of considering a large number of objectives on decision difficulty is relevant even outside the MCDM/EMO communities. Consumers' choices of products or services, for instance, are typically based on attributes that distinguish these products or services from others and are in some ways linked to the consumers' objectives (e.g., Fasolo, McClelland, and Todd 2007). Online shopping sites (such as PriceRunner.com) seek to help consumers find the best product by providing information on dozens of attributes. Yet, the abundance of information may in fact make the task of comparing products more instead of less difficult (Malhotra, 1982). In the domain of consumer choice, a practical problem setting analogous to multicriteria decision making or multiobjective optimization is choosing between bundles of products or services, where products/services correspond to criteria/objectives. Marketers may be tempted to introduce bundles with increasingly complex ranges of products or service features to increase their competitiveness and attract new customers (Agarwal & Chatterjee, 2003; Miao & Jayakar, 2014). Yet, such efforts may turn out to be counterproductive, if choices between complex bundles become overly difficult for consumers (Agarwal & Chatterjee, 2003; Iyengar, Ansari, & Gupta, 2007).

The above observations generate several interesting research questions:

1. Does increasing the number of objectives lead to higher cognitive burden for the decision-maker?
2. How does increasing the number of objectives affect the way in which decision-makers process and use information?
3. How does increasing the number of objectives affect decision quality?
4. Does a specific number of objectives exist beyond which it is "too" burdensome for humans to make high-quality decisions?

These questions are equally relevant for MCDM and EMO scholars as well as practitioners attempting to formulate and solve multiobjective decision-making problems. In particular, a push for considering more and more objectives may be ultimately fruitless, if it is found that the resulting increase in cognitive burden leads to decision-makers ignoring much of the information and not being able to make high-quality choices.

To find answers to our research questions, we have conducted a comprehensive laboratory study with 50 human participants making altogether 5760 decisions.[2] In this experiment, we approached the difficulty of decision-making in terms of the limitations of human cognitive capacity to process a lot of information (DeLeeuw & Mayer, 2008; Gigerenzer & Todd, 1999; Miller, 1956; Simon, 1955; Sweller, 1988; Cowan, 2010, 2012; Paas, Tuovinen, Tabbers, & Van Gerven, 2016). The task of the participants in the experiment was to make binary (pairwise) comparisons between two multiobjective decision alternatives represented by bundles consisting of everyday consumer products.[3] We systematically varied the number of product types (dimensions) in each comparison between 2 and 10. To analyze the cognitive burden caused by adding dimensions to the binary comparisons we use mixed methods, including psychophysiological (notably eye tracking), behavioral, and self-report measures for all central constructs.

When it comes to decision quality, various definitions have been presented in the literature. Hammond (1996, 2007), for instance, makes a distinction between coherence and correspondence theories in decision making. Coherence theories emphasize rationality and internal consistency and are based on normative standards of logic, whereas correspondence theories focus on the empirical accuracy of judgments and decisions. Arkes, Gigerenzer, and Hertwig (2016) argue that in real-life decision settings, coherence-based norms are of limited value for evaluating decision quality, because computational intractability and conflicting goals can make coherence unattainable, and because incoherence can in fact be a sign of adaptive learning. The norms of rationality have also been criticized by March (1991, 2006), particularly in the context of organizational decision-making.

In examining the performance of simple heuristics and complex decision models, Katsikopoulos, Durbach, and Stewart (2018) distinguish between repeated operational decisions (including tasks of inference and forecasting) and strategic one-off decisions. They note that in inference and forecasting tasks, decision quality can be assessed with respect to some objective standard, whereas this is not possible in strategic decision-making tasks that necessarily involve subjective preferences. In the literature examining the performance of simple heuristics in multiobjective decision-making tasks, decision quality is often measured through the proportion of value lost due to the use of a heuristic instead of a more complex preference model such as a multicriteria value/utility function (Barron, 1987; Fasolo et al., 2007) or the share of decision settings in which a heuristic would lead to choosing the same decision alternative as the more complex model (e.g., Methling, Abdeen, & von Nitzsch, 2022). Yet, these approaches assume that the more complex model correctly represents the decision-maker's preferences. Proportion of decision settings in which a non-dominated alternative was chosen is an attractive measure for decision quality because it is independent of the decision-maker's preferences (Baucells, Carrasco, & Hogarth, 2008; Kourouxous & Bauer, 2019; Pande, Papamichail, & Kawalek, 2021). Yet, such a measure cannot be applied when all decision alternatives that are being considered are non-dominated — which is the case in our experiment.

In this paper, we study the impact of dimension on decision quality through measuring the consistency of the participants' subjective choices. In particular, we examine the share of cases in which a participant makes identical choices in identical decision settings (duplicate

[2] We conducted our study in Professor Niklas Ravaja's behavioral laboratory at the Department of Psychology and Logopedics, University of Helsinki. Niklas Ravaja's research staff has extensive experience in conducting psychophysiological measurements and analyzing such data. Our sample size is relatively common for a within-subjects study utilizing psychophysiological measures; see, e.g. Bellemare, Bissonnette, and Kröger (2014), Chikhi, Matton, and Blanchet (2022) and Li, Luh, and Chen (2024).

[3] Although several preference elicitation mechanisms have been presented in the literature, we focus on binary (or pairwise) comparisons of multidimensional vectors (of objectives) due to their fundamental nature and widespread use.

consistency) and whether their choices coincide with their preferences regarding different product types (preference consistency). Duplicate consistency is a central tenet of rational choice theory (Arrow, 1982; Luce & Raiffa, 1957), and can therefore be seen as a necessary (albeit not sufficient) measure for decision quality. We complement duplicate consistency with preference consistency to utilize the information that we gathered regarding our subjects' preferences for different product types. In computing the preference consistency measure, we make different assumptions about the ways in which the participants' preferences are inferred and aggregated. The duplicate consistency measure avoids making such assumptions but, on the other hand, does not utilize any of the available preference information. While consistency is but one aspect of decision quality, lower decision consistency can be seen as an indication of lower decision quality, especially if the evidence provided by the two measures is mutually supportive.

The rest of this paper is organized as follows. Section 2 presents a brief overview of existing psychological studies concerning psychological burden and information overload. In Section 3, we provide a detailed account of the experiment. Section 4 introduces the measures used for analyzing cognitive burden and its impacts on behavior and decision consistency. Statistical analyses on these measures are presented in Section 5, and Section 6 concludes.

## 2. Theoretical background and hypotheses

The effects of too much information have been studied in the information overload literature (see Roetzel 2019). Information overload (sometimes used synonymously with cognitive overload) is a concept that has no single definition, but generally describes a mental state that occurs when the amount of information exceeds the individual's capacity to process it and that leads to or is associated with higher negative affect (or emotional feeling) and a drop in decision speed and quality (e.g., Bawden & Robinson, 2009; Rutkowski & Saunders, 2018). In practice, the concepts of overload imply that the cognitive work happens unrestricted until some threshold is reached, at which point the overload begins a system failure resulting in problems like negative feelings and avoidance behaviors. Information overload may lead to the adoption of different strategies to be able to carry on with the task. For instance, Peng, Xu, and Huang (2021) found that in response to inducing information overload, the participants allocated fewer mental resources to the task rather than more. Information overload has also been found to result in the adoption of different kinds of stopping rules for information gathering (Browne, Pitts, & Wetherbe, 2007; Pennington & Kelton, 2016) as well as filtering and withdrawing from too much information (Savolainen, 2007).

In terms of cognition, information overload can be seen as a result of the limitations of working memory, which is studied as cognitive load (CL) in psychology (Paas, Van Gog, & Sweller, 2010). Popular neuroscientific theories suggest that working memory is the process that integrates information from different brain processes into one stream that consciousness can use in decision making (Dehaene, 2014). Notably, all conscious, or effortful, decisions (see "System 2", Kahneman, 2011) need to go through this one bottleneck process, which has natural limitations to its capacity. Different attempts to pinpoint a universal number for this limit have produced different figures (e.g. 7+/−2 units, by Miller, 1956; 3–4 units, by Vogel, Woodman, & Luck, 2001; 3–5 units, by Cowan, 2010), but each of these figures is quite low. Information or cognitive overload thus could be understood to occur when the working memory resources needed are greater than available resources.

In this paper, we are interested in studying *cognitive burden*, by which we refer specifically to the ways in which cognitive load affects behavior. A strictly rational agent would utilize working memory regardless of its load to the maximum according to the needs of the task. However, CL, as demonstrated by the phenomenon of information overload, is uncomfortable — the human brain prefers to not do the work if it can avoid it (Kahneman, 2011). We consider cognitive burden specifically in the context of how CL influences decision-making because of the human mind's natural avoidance of effort and discomfort (see Feldon, Callan, Juth, & Jeong, 2019, for a similar view).

From a psychological point of view, discomfort is a form of affect, and its influence on behavior is conceptualized as motivation (Eder, Elliot, & Harmon-Jones, 2013; Pezzulo, Rigoli, & Friston, 2015). Affect is considered as a signal that the current goals are being fulfilled (positive affect) or not (negative affect) by the current activity (Barrett, 2017; Barrett & Finlay, 2018). A positive affect signals that the person should continue whatever they were doing. A negative affect, however, should lead to changes in behavior (do the task in another way to make reaching the task goals more likely), in goals (change goals to make meeting them easier), or both. From the affective-motivational point of view, then, cognitive burden would involve a negative affective reaction to the assessment that the time and effort required to process the large amount of information is going to be considerable. This conflicts with biological goals to minimize discomfort and conserve energy and may conflict with or be supported by goals from the personal context (e.g., to strive for a bonus or to impress someone). One way to resolve the affective-motivational conflict is the reprioritization of goals: the decision-maker (or more accurately, their brain, nonconsciously) concludes that it is more important to save effort and/or time than carry out the task optimally, if the task is completed at least with some accuracy.

The crucial difference between information/cognitive overload and our affective-motivational conceptualization of cognitive burden is that cognitive burden is not a separate state that only occurs when capacity is exceeded; instead, cognitive burden is the characteristic of all CL that is experienced as negative affect and that motivates the individual to avoid the effort. In other words, cognitive work is constantly being resisted from the beginning, and this resistance is what we call cognitive burden. Moreover, while the negative affect is explicitly associated with behavior change, to which extent this is realized depends on whether the negative affect is stronger than the counteracting positive affects that motivate the individual to spend the effort in order to reach the goals. That is, cognitive burden does not reflect a failure of the human cognitive system and the resulting unintended problems, but its operation as intended by active adaptation. In a sense, information/cognitive overload can be understood as a concept describing the aggregate effects of high burden in a sample (usually studied after the fact), as opposed to a more precise phenomenon of cognitive burden that can be measured while it is happening.

In this paper, we study cognitive burden and dynamic adaptation in the context of making pairwise choices between Pareto-optimal, multiobjective decision alternatives. As the processes are essentially psychological, we use psychological methodology and experimentation. In our context, a larger number of objectives is likely to increase the CL associated with comparing multiobjective alternatives. This motivates us to investigate the following hypothesis related to our first research question ("Does increasing the number of objectives lead to higher cognitive burden for the decision-maker?"):

*H1: Cognitive burden increases with the number of objectives.*

Choices between two Pareto optimal solutions, in particular, depend on the decision-makers' subjective preferences, which makes these situations susceptible to goal reprioritization when the assessment of required effort becomes too high. Because a preference task does not force the load on the participants, we predicted that they could employ dynamic adaptation — that is, switch goals preemptively according to expected load and minimize the associated negative affect (Barrett, 2017; Friston, 2010; Pezzulo et al., 2015). Such adaptation could be manifested by focusing only on a limited subset of objectives (Fasolo et al., 2007; Katsikopoulos et al., 2018; Pande et al., 2021). To study whether this happens, we investigate the following hypothesis related to our second research question ("How does increasing the number of objectives affect the way in which decision-makers process and use information?"):

1. A Bookbeat audiobook streaming service for 20 hours,
2. McDonald's gift card for one meal,
3. A 6-pack of Coca-Cola,
4. A design shopping bag by Finlayson Co,
5. One Finnkino movie theater ticket,
6. Credits worth five euros to the Foodora food delivery platform,
7. 450 g pack of Pelican Rouge fair trade coffee,
8. Kotipizza restaurant chain gift card for a pizza of participant's choice,
9. Moleskine notebook,
10. A box of Fazer chocolate bars,
11. A tube of Pepsodent toothpaste,
12. 1-month subscription to a Spotify premium music streaming service,
13. Suomalainen bookstore gift card worth 10 euros,
14. A five-euro gift card to Teekauppa online tea shop,
15. A 10-euro gift card to Mall of Tripla stores.

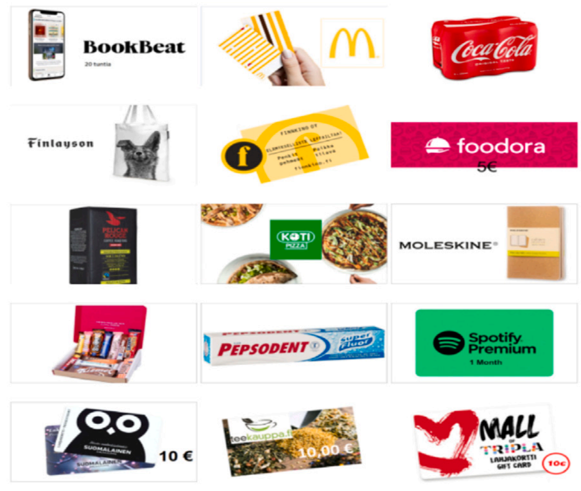


**Fig. 1.** Descriptions and images of $J = 15$ product types.

*H2: When the number of objectives is high, decision-makers adapt to anticipated negative affect by considering a limited number of objectives only.*

The increasing cognitive burden and goal reprioritization may result in lower decision quality which, in this paper, is examined through the concept of consistency. To address our third research question ("How does increasing the number of objectives affect decision quality?"), we study the following hypothesis:

*H3: Decision consistency decreases with the number of objectives.*

For the fourth research question ("Does a specific number of objectives exist beyond which it is "too" burdensome for humans to make high-quality decisions?"), we anticipate that the result may depend on both the individual and the task at hand. Thus, we approach this question in an explorative manner without specifying hypotheses.

## 3. Experimental setup

Our study was reviewed and approved by the Ethics Review Board in the Humanities and Social and Behavioral Sciences of the University of Helsinki. All participants gave informed consent prior to the experiment and were debriefed after the experiment. In this section, we describe the experiment in detail.

### 3.1. Experimental subjects

The experimental subjects were 50 adult volunteers recruited via the University of Helsinki psychology student organization email lists. Care was taken to recruit as close to an equal number of male and female participants as possible. Due to the requirements of the physiological measures, we prescreened the participants to be right-handed people, who had normal eyesight and hearing, and who were physically and mentally healthy. We had trouble with the measuring equipment with two participants, so in many (but not all) analyses, the sample is the remaining 48 participants.

### 3.2. Study design and procedure

In the experiment, the participants faced tasks of choosing between two bundles $\boldsymbol{x} = (x_1, \dots, x_J)$ and $\boldsymbol{y} = (y_1, \dots, y_J)$ of everyday products, where $x_j, y_j \in \mathbb{Z}$ are the numbers of copies of product type $j \in \{1, \dots, J\}$ in bundles $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. These bundles consisted of altogether $J = 15$ product types, which are specified in Fig. 1. Some of these product types are specific to the Finnish consumer market. The aim was to choose product types that the participants would appreciate in that more copies would be preferred to fewer copies for each product type. While a choice between bundles may be a fairly non-typical MCDM/EMO task, bundles can be thought of as multiobjective[4] decision alternatives, where the different product types represent different objectives (cf. Korhonen and Wallenius 2020, Korhonen, Wallenius, Genc, and Xu 2021 for a similar representation). We chose to use bundles as alternatives in our experiment because this was something that the subjects (psychology students) were familiar with and, due to the choice of product types, could relate to in their everyday lives. These aspects were found to be important because of the length and tediousness of the experiment. As a further attempt to avoid making the experiment overly complicated, we focused on pairwise comparisons between bundles. While the choice in most MCDM/EMO models is ultimately made among many Pareto-optimal alternatives (instead of only two), pairwise comparisons between multiobjective decision alternatives are a fundamental and commonly used tool in eliciting decision-makers' preferences to support the selection of a single most-preferred Pareto-optimal solution (Figueira, Greco, Mousseau, & Słowiński, 2008; Kadziński, Greco, & Słowiński, 2012; Korhonen, Wallenius, & Zionts, 1984; Phelps & Köksalan, 2003).

In each choice setting, both bundles $\boldsymbol{x}$ and $\boldsymbol{y}$ contained the same number of product types. The number of product types in a bundle is referred to as the *dimension of the bundle*, formally defined as follows.

**Definition 1.** Consider bundle $\boldsymbol{x} = (x_1, \dots, x_J) \in \mathbb{Z}^J$. The dimension of the bundle, denoted by $\mathrm{Dim}(\boldsymbol{x})$, is defined as the number of different product types included in the bundle, i.e.,

$$\mathrm{Dim}(\boldsymbol{x}) = \left| \left\{ j \in \{1, \dots, J\} \middle| x_j > 0 \right\} \right|.$$

The dimension of the bundles in a given choice setting is comparable to the number of objectives in a multiobjective decision-making problem, although the two concepts are not exactly the same.[5] The number of copies of each product type in each bundle (i.e., the values of $x_j$ and $y_j$) ranged between 0 and 4. The bundles were created so

[4] The terms 'multicriteria' or 'multiattribute' could be used here as well.

[5] Consider, for instance, a choice between bundles $\boldsymbol{x} = (1, 0, 3)$ and $\boldsymbol{y} = (0, 1, 3)$. The dimension of the bundles in this choice setting is $\mathrm{Dim}(\boldsymbol{x}) = \mathrm{Dim}(\boldsymbol{y}) = 2$, while in the context of multiobjective optimization, the number of objectives is 3.

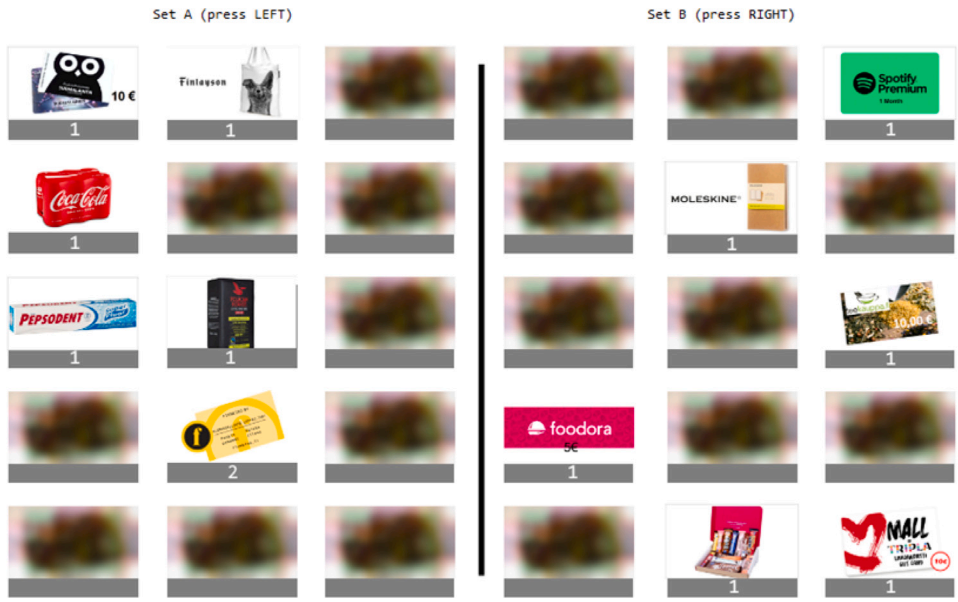


**Fig. 2.** A recreation of a choice setting presented to the participants with $\mathrm{Dim}(\boldsymbol{x}) = \mathrm{Dim}(\boldsymbol{y}) = 6$.

that neither bundle dominated the other (i.e., neither bundle contained a larger or equal number of copies of all product types and a strictly larger number of copies of at least one product type compared to the other bundle). That is, choices would be made between bundles such as $\boldsymbol{x}$: (2 movie tickets, 1 box of chocolate bars, 2 shopping bags) and $\boldsymbol{y}$: (1 movie ticket, 2 boxes of chocolate bars, 1 bookstore gift card). Thus, when choosing between $\boldsymbol{x}$ and $\boldsymbol{y}$, a subject would have to consider whether they would be, e.g., willing to sacrifice one movie ticket to get an additional box of chocolates, and to give up two shopping bags to get one bookstore gift card. Thus, a choice between bundles would entail making trade-offs between different product types, which is a key feature of multiobjective decision-making problems. Moreover, the bundles were set to have roughly equal monetary (retail) values so that one bundle would not be objectively superior to the other.

Each choice setting was presented to the participants through two grids as illustrated in Fig. 2. The left-hand-side grid (titled “Set A”) corresponds to bundle $\boldsymbol{x}$, and the right-hand-side grid (titled “Set B”) to bundle $\boldsymbol{y}$. Each grid contained the pictures of the product types included in the bundle so that the number of copies of a given product type was shown below the corresponding picture. In the example of Fig. 2, both bundles contain six product types (i.e., $\mathrm{Dim}(\boldsymbol{x}) = \mathrm{Dim}(\boldsymbol{y}) = 6$). In this example, bundle $\boldsymbol{x}$ includes, among others, a single six-pack of Coca-Cola cans and two movie tickets. The locations of different product types in the grid were randomized but held constant across all trials for a given participant. Empty slots in the grids corresponding to product types not included in the bundle were filled with visual noise created by combining all the product images together. Above each grid, an instruction text was shown to inform the participant about which key they should press to choose the bundle.

Before starting the experiment, the subjects practiced the procedure with three example choices that were not part of the actual experiment. They were also shown the full list of the 15 products used in the experiment and were asked to rate their preference for each product on a scale from 1 to 9. The rating by subject $i$ for product $j$ was denoted by $r_{ij}$. At this point, we did not provide information on the prices of the product types in fear of such information acting as an anchor for the subjective product ratings. Still, because of the familiarity of the product types for our subjects, they may have had at least a rough idea about these prices anyway. In principle, a positive correlation between prices and product ratings would not be an issue — it is quite natural that more expensive product types would also be subjectively preferred. Nevertheless, the Spearman correlation between the products’ monetary values and the subjects’ ratings was only 0.235. The subjects were also asked to report how tired, distracted, or nervous they felt (questions A1–A3 in Table A.1 in Appendix A). Before running the experiment in full, we tested five pilot participants in order to ensure that the experiment ran as intended and that the procedure was not too uncomfortable for the participants. The experiment took 2–3 h per participant, including the set-up and debriefing time.

The experiment was run as a block design with five decision-making conditions corresponding to different dimensions $\mathrm{Dim}(\boldsymbol{x}) = \mathrm{Dim}(\boldsymbol{y}) \in \{2, 4, 6, 8, 10\}$ (i.e., numbers of product types in each bundle). In total, there were 10 blocks, two for each dimension. Each block contained 12 trials, 10 with unique bundles, and 2 repetitions of already presented bundles, where the positions of bundles $\boldsymbol{x}$ and $\boldsymbol{y}$ had been reversed. After each block of 12 trials, self-report measures were administered (see Table A.2 in Appendix A for the list of questions for these self-report measures), and the participants could take a small break if they wanted.

In the post-experiment phase, the participants filled in trait questionnaires.[6] Then, the participants were asked to answer some post-experiment self-report questions about how they experienced the choices and made them, what they thought were the sources of difficulty, and what kinds of strategies they used (see questions C1–C27 in Table A.3 in Appendix A).

### 3.3. Compensation

The participants were compensated with a lottery scheme, where 90% of participants received a bundle (of their choice) consisting of the products used in the experiment worth up to 25 Euros, and 10% of participants received a bundle worth up to 90 Euros. For practical reasons, the complete reward scheme was not disclosed in detail before the experiment. Rather, the participants were told that they would get one of the product bundles they chose during the experiment to create a motivation to choose the bundle they actually preferred in each choice setting. After the experiment, the participants were, however, given a choice: to wait for the analysis of which bundles they actually chose, or to freely choose products up to the euro limit right away. None of the participants chose to wait.

[6] Since the study of individual differences is not our focus in the current paper, we do not provide the details of the trait questionnaires.

## 4. Measuring cognitive burden and its effects on behavior and decision consistency

We examined the effects of cognitive burden in two steps: first, whether we could detect CL and affective and other responses indicating burden; second, whether differences in burden changed behavior and decision consistency.

### 4.1. Cognitive burden

Cognitive burden was examined through five associated measures that were intended to examine whether CL occurred (CL, blink rate), and whether the CL that did occur was associated with expected experiences (affect, perceived task complexity, perceived information overload). Following common practice in the literature (DeLeeuw & Mayer, 2008; Paas et al., 2016), we measured **CL** with self-report measures regarding the effort needed and the difficulty of the task. In particular, we used measure

$$\text{Self-reported CL} = \frac{1}{2}\,(\text{B1} + \text{B2})\,, \tag{1}$$

where B1 and B2 are the responses to questions posed to the subjects after each block regarding the level of effort and difficulty of the task measured on a nine-point scale (see Table A.2 in Appendix A for details). Self-reports were only recorded at block level, since asking the questions after each trial would have made the experiment very tiring and cumbersome.

In addition to self-report measures that can sometimes be unreliable, we also used methods not reliant on the participants' voluntary and conscious answers. As a physiological, nonvoluntary indication of CL, we studied the subjects' **blink rate** (blinks per second or minute). The absolute values of blink rate may vary between individuals, for instance due to size differences in eye muscles that are irrelevant for the psychological processes we intend to measure. This means that the absolute values are not of importance, but only the changes between conditions are (Cacioppo, Tassinary, Berntson, et al., 2007). Blink rate decreases in response to the need to process visual stimuli more effectively (Cardona & Quevedo, 2014; Paas et al., 2010). Specifically in short (2-20s), visual, moderate-load trials where the eyes are not fixed to a single point, increase in cognitive processing is reflected by decrease in blink rate due to attentional focus maximizing information extraction (after information search; Skaramagkas et al., 2021; Wilbanks, Aroke, & Dudding, 2021). In other words, a decreasing blink rate is associated with higher CL. The more common eye-related measure of CL, pupil size, could not be used due to the amount of eye movements that were essential for comparing the two bundles.

To measure self-reported affect, the subjects were asked after each block to respond to verbal questions (following the PANAS self-report scales; Watson and Clark 1994[7]) regarding their affect and motivation (see Table A.2 in Appendix A). In particular, they were asked to specify the extent to which they felt uncomfortable (B4), frustrated (B5), bored (B6), and not wanting to do the task (B14). The responses to these questions were used to define a measure for **self-reported negative affect** as follows:

$$\text{Self-reported negative affect} = \frac{1}{4}\,(\text{B4} + \text{B5} + \text{B6} + \text{B14})\,. \tag{2}$$

In addition to the above measure, we used SAM (Self-Assessment Manikin) rating scales (Bradley & Lang, 1994) to measure valence, arousal, and dominance (questions B8–B10 in Table A.2 in Appendix A). SAMs are non-verbal (picture-based) measures of affect that sometimes reflect the physical state of emotion better than verbal measures. Valence is a bipolar measure of affect – when the affect is more positive, it is also less negative – and it is accompanied by arousal, or intensity. Dominance is the feeling of being in control.[8] Results on these non-verbal measures are used in this paper to find potentially corroborating evidence for the verbal measures.

In addition to specific measures of CL, we also used separate self-report measures for experienced **complexity and information overload**. According to Roetzel (2019), information overload (including perceived complexity) has no standard measure, and is commonly assessed with ad hoc measures. Here, we use the following single-item measures:

$$\text{Self-reported perceived complexity} = \text{B11} \tag{3}$$

$$\text{Self-reported information overload} = \text{B12}, \tag{4}$$

where B11 and B12 are the responses to questions posed to the subjects after each block on whether the task seemed complex, and whether the participants felt overwhelmed by the task (Rutkowski & Saunders, 2010) measured on a nine-point Likert scale indicating agreement (see Table A.2 in Appendix A for details).

### 4.2. Behavior

The measures of behavior were direct task performance metrics: response time and eye-tracking measures. Further evidence on behavioral changes was gathered through post-experiment questionnaires where the participants reported their own views about their behavior (see questions C1–C27 in Table A.3 in Appendix A).

**Response time (RT)** was measured (by the stimulus software, in milliseconds) from the first appearance of the choice setting on the screen to the choice key being pressed by the participant. RT is a direct measure of effort spent on the task — the more work the participant puts in, the more time it takes. DeLeeuw and Mayer (2008) mention RT as a typical and valid measure of (a facet of) CL.

The eye-tracking measures used in this study reflect how the participants process the information. Human eyes operate by focusing on a single point of interest (fixation) at a time, with sudden jumps (saccades) between them (Rayner, 1998). Information intake occurs during the fixation, while during a saccade the brain does not update information from the environment. The number of fixations and their duration are reported to increase with task difficulty (Perkhofer & Lehner, 2019). This is because more information to parse through requires more fixations and more complex information requires longer fixations, and when the fixations are on average closer to each other, the saccades get shorter and quicker.

In vision research, fixations can be divided to short and long ones, where short fixations (around 100–200 ms) are commonly considered as search behavior — the gaze briefly visits a location, but the brain decides the location did not contain relevant information, so it quickly moves on to the next location (Negi & Ritayan, 2020; Rayner, 1998). This processing that is only concerned with finding the information for the decision-making (information intake) can be contrasted with the process of decision-making itself, where the participant does the cognitive work to compare values between the bundles and ultimately come to a decision. When the brain is engaged with this deeper processing, additional intake of information via continuous saccades is suppressed, resulting in longer fixations. In this study, two eye-tracking measures are used: **total fixation duration**, indicating the sum of information intake and cognitive work; and **total long fixation duration**, indicating cognitive work without information intake. Different studies have used various cutoffs for categorizing fixations into short and long

[7] There is no directly suitable instrument for the negative affect related to the avoidance of effort. We used the general question format of PANAS-X but changed the wordings to expressions from Frustration and Boredom scales (Harrington, 2005; Struk, Carriere, Cheyne, & Danckert, 2016) and ad-hoc items that we considered more appropriate.

[8] Note that despite the term being the same, this concept of (emotional) dominance is different from the one used to compare multiobjective decision alternatives.

ones (Negi & Ritayan, 2020); here, following Glöckner and Herbold (2011), a fixation is counted as long if it lasts more than 250 ms.

Note, however, that it is not the difficulty of the task per se that affects these measures, but rather the increased effort to meet the demands of the task. The effort depends on the person and their motivation to increase the effort. If the duration of fixations increases linearly with dimension (i.e., the number of products in the bundle), it implies a strategy that always samples a proportional amount of information, rather than a strategy dependent on particular products. If the duration of fixations does not increase with dimension at all, it implies that the strategy depends on a constant amount of information that is easily accessible due to fixed product locations. A bent graph indicates that the strategy considers a fixed maximum number of pieces of information, and consistently ignores the rest of the information.

### 4.3. Decision consistency

Decision consistency is difficult to define when the task involves expressing one's preferences, a subjective task goal by definition. We used two measures of consistency: duplicate and preference consistency.

**Duplicate consistency** is defined as the share of identical choices when duplicate product bundles were presented within the same block. The two duplicate product bundles were presented in reverse order the second time, to avoid easy detection by participants. Since each dimension level had two blocks, the chance for randomly picked bundles on a given dimension level to be completely duplicate consistent was $0.5^4 = 0.0625$. Assuming that the criteria for making a decision do not change within the block, a rational decision-maker should make the same choice when faced with the same choice setting, regardless of their individual preferences.

**Preference consistency** is a measure that aims to utilize information regarding the participants' individual preferences. Under the axioms by Krantz, Luce, Suppes, and Tversky (1971), a decision-maker's preferences regarding different bundles can be captured by a measurable value function $V : \mathbb{Z}^J \to \mathbb{R}$. If a value function $V_i$ could be determined for each subject $i$ in our experimental study, then we could observe whether their choices were consistent with their preferences (i.e., whether they always chose the bundle that resulted in a higher value). Yet, the elicitation of subject-specific value functions that would capture interactions between all product types would have been very difficult, time-consuming, and out of scope for the current paper. Instead, we use simple linear-additive value functions with linear product-specific value functions as proxies for measuring the consistency of the subjects' choices with their preferences:

$$V_i(\boldsymbol{x}) = \sum_{j=1}^{J} w_{ij} x_j, \tag{5}$$

where $w_{ij}$ is the (non-normalized) weight attached by subject $i$ to product type $j$ reflecting the subject-specific value of increasing product type $j$ from its worst level ($x_j = 0$) to its best level ($x_j = 4$). Thus, weight ratios $w_{ij}/w_{ij'}$ represent tradeoffs between different product types $j, j'$ for subject $i$.

These simple value functions are based on several unverified preference assumptions (namely, measurable and linear product-specific value functions as well as mutual preference independence and difference independence between different product types; see Dyer and Sarin 1979), whereby they cannot be said to accurately represent our subjects' preferences. Yet, they serve a useful proxies for studying whether the consistency of the subjects' choices with their preferences is affected by the dimension of the choice setting. In particular, preference consistency is defined as the share of choice settings in which the choice made by the participant coincided with the one that maximized the proxy value function for this participant as in Eq. (5). It should be noted that the suitability of this measure for decision consistency depends on not only the aforementioned preference assumptions, but also on the static nature of the subjects' preferences (involving no learning).

We compute preference consistencies for four different choices of $w_{ij}$ (weight attached by subject $i$ to product type $j$). First, we use the subjects' self-reported ratings for different product types as weights, i.e., $w_{ij} = r_{ij}$. These ratings were, however, given by using a limited, discrete scale 1–9, which may not have been interpreted by our subjects as a cardinal scale that would capture tradeoffs between the product types. That is, a subject giving ratings $r_{ij} = 8$ and $r_{ij'} = 2$ may not have meant that one copy of product type $j$ would compensate for $r_{ij}/r_{ij'} = 4$ copies of product type $j'$. Instead, the scale may have been used in an ordinal sense. Surrogate weights have been proposed as a means to convert ordinal preference statements into cardinal ones (Danielson & Ekenberg, 2017). Thus, in addition to the self-reported ratings, we compute preference consistencies for three commonly used sets of surrogate weights: rank sum weights ($w_{ij} = w_{ij}^{RS}$), rank reciprocal weights ($w_{ij} = w_{ij}^{RR}$), and rank-order centroid weights ($w_{ij} = w_{ij}^{ROC}$). Technically, these weights are computed from the self-reported ratings as follows:

$$\text{Rank sum weights:} \quad w_{ij}^{RS} = \frac{r_{ij}}{\sum_{\ell=1}^{J} r_{i\ell}}, \tag{6}$$

$$\text{Rank reciprocal weights:} \quad w_{ij}^{RR} = \frac{1/(J+1-r_{ij})}{\sum_{\ell=1}^{J}(1/(J+1-r_{i\ell}))}, \tag{7}$$

$$\text{Rank-order centroid weights:} \quad w_{ij}^{ROC} = \frac{1}{J}\sum_{\ell=j}^{J}\frac{1}{J+1-r_{i\ell}}. \tag{8}$$

## 5. Statistical analyses on the impact of dimension on cognitive burden

In this section, we present the results of statistical analyses on the impact of dimension (i.e., the number of products in each bundle) on cognitive burden as well as behavioral changes and decision quality. These results are used to answer the research questions posed in Section 1. The analyses are based on linear mixed models that were fitted for the measures described in the previous section with dimension as the independent variable. Linear mixed models were used because of the within-subjects design of the study, which creates dependencies within individuals, and requires the control of these dependencies with random effects. In cases where the outcome variable was not continuous (duplicate consistency and preference consistency), logistic mixed models were used. Before running the analyses, the mixed model assumptions were checked and when necessary, the data was transformed with Box–Cox transformation that helps stabilize variance and make the data more normal distribution-like (Box & Cox, 1964). The values reported in this paper have been transformed back to the original units. Due to issues with heteroscedasticity and non-normality, this transformation was performed on blink rate, reaction time, and eye fixation measures. In case of self-reports, one or two outliers were removed when suggested by the diagnostic analysis.

Models with two levels (trial, participant) were run for each primary trial-level measure (i.e., RT, eye fixation measures, and the preference consistency measure), and models with one level (participant) for the self-reported block-level measures (i.e., CL, negative affect, perceived complexity, and information overload) and the duplicate consistency measure. The models were fitted in a two-stage stepwise manner, comparing the goodness of fit in terms of Akaike Information Criterion and Bayesian Information Criterion, which penalize model complexity less and more severely, respectively. In the first stage, we compared three models: a random intercept model, a random intercept and random slopes model, and a random intercept and random slopes with the trial (or block, in one-level models) count as an additional predictor. Using the best model found as the base case, in the second stage we built three piecewise linear models with a bend at a dimension of 4, 6, and 8, respectively. Then, we chose the best one among the four competing

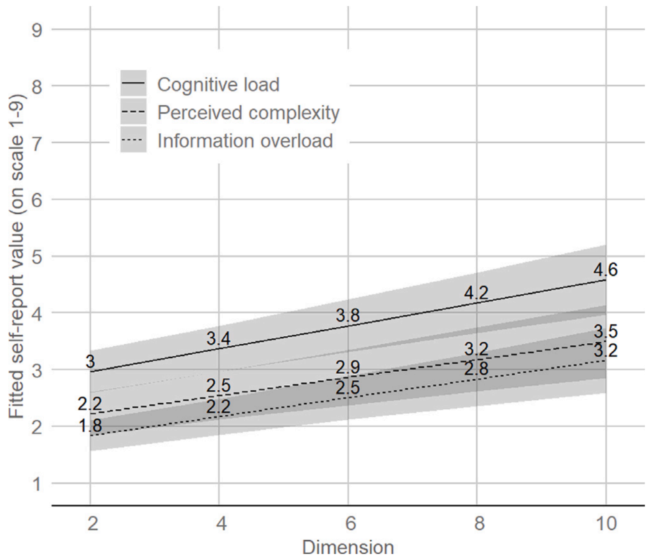


**Fig. 3.** Self-reported experience of CL, task complexity, and information overload in relation to the dimension of the bundles in the choice setting. Values at fixed dimensions are shown on top of the lines.

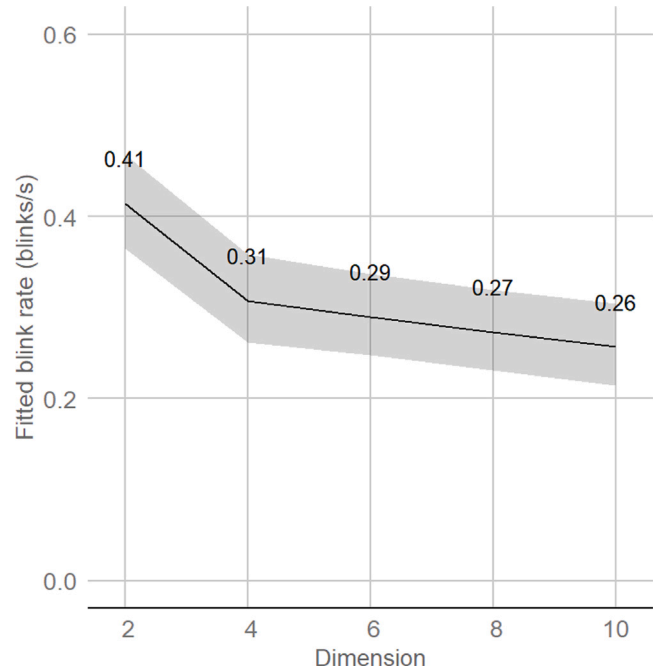


**Fig. 4.** Blink rate (times per second), indicating CL, in relation to the dimension of the bundles in the choice setting. Lower blink rate is associated with higher CL.

models (the linear model and the three piecewise linear models). The tested models, diagnostic analyses, and their results for each measure are detailed in the Supplementary material.

Before running our statistical analyses, we screened the participants for obvious abnormal response patterns (e.g., making left choice for all comparisons). No such patterns were found. Standard $p < 0.05$ criteria were used with one-tailed tests. Analyses were run in R versions 4.1.1 and 4.3.1 (R Core Team, 2015).

### 5.1. Did higher dimension add to the cognitive burden?

Fig. 3 shows the fitted regression lines for self-reported CL, perceived task complexity, and information overload (cf. Eqs. (1), (3) and (4)) as functions of the dimension of each bundle in the choice setting. The gray areas describe the 95% confidence intervals for the fitted values. Based on this figure, all three measures increase with dimension. This indicates that the participants felt the task was more complex, demanded more effort, and caused more feelings of being overwhelmed when the bundles contained more product types. Still, the overall level of all three measures was quite low, suggesting that these experiences were not very strong. This may be an indication of the participants regulating the burden themselves through setting their own tolerance level for acceptable discomfort.

Fig. 4 shows the regression line for blink rate (blinks per second), a measure that is commonly used as a proxy for CL. The absolute values of blink rate depend on the individual and are of no importance, only the changes between conditions are. We observed that the average blink rate decreased with dimension, supporting the self-reports that CL was increasing as a function of dimension. This decrease seems to be largest when the dimension increases from two to four — a feature that was not apparent in the self-reports.

Next, we studied the dependence of self-reported negative affect on the dimension of the bundles in the choice setting (cf. Eq. (2)). The theoretical view predicted that higher CL would cause negative affect. Based on our statistical analyses, the self-reported negative affect increased only very slightly with dimension. However, when we ran a mixed model for negative affect with dimension as an independent variable and CL as a covariate, we found that negative affect was predicted by CL but not by dimension (see Table 1). In other words, while a higher dimension was associated with more negative feelings,

**Table 1**
Fixed effects of the mixed model for self-reported negative affect.

| Variable | $\beta$ estimate (SE) | t-value | p-value |
|---|---|---|---|
| Intercept | 1.029 (0.151) | 6.837 | <0.0001 |
| Dimension | 0.030 (0.016) | 1.950 | 0.0559 |
| Block | 0.099 (0.011) | 9.353 | <0.0001 |
| CL | 0.212 (0.028) | 7.640 | <0.0001 |

those feelings were related to the cognitive burden that increased with the dimension, rather than the dimension itself. In addition, negative feelings were associated with block — that is, the longer the participant had done the task, the more negative they felt.

We found corroborating evidence from the non-verbal Self-Assessment Manikin measures through fitting regression models for the responses to questions B8–B10 (see Table A.2 in Appendix A). These models (reported in detail in the Supplementary material) suggest that similarly to verbal measures, affect turns a bit more negative with higher dimension, but that this effect is much better attributable to CL than dimension as such (cf. Table 1). Regarding valence (i.e., question B8), a piecewise linear model indicates that a choice setting with dimension 2 (i.e., in which both bundles only contain two project types) is associated with slightly higher (i.e., more positive, less negative) valence than all the other dimension levels. This may be because it is the only dimension level where all the products present on the screen (representing at most four product types) would easily fit in the working memory. In addition, we found that while differences in emotional arousal or dominance do not seem to depend on dimension alone, they are linearly associated with CL (higher CL associated with higher arousal and lower feeling of dominance; the model for dominance, however, was found to be unreliable). This indicates that, as expected by our conceptualization of cognitive burden, decisions where the effort required was the highest were the most unpleasant to experience, although the differences were quite small.

The above findings were further supported by responses to the post-experiment questionnaires (see Table A.3 in Appendix A). The participants, on average, agreed (on a scale from 1 = Strongly disagree to 7 = Strongly agree) that: “Looking at and keeping in mind all the

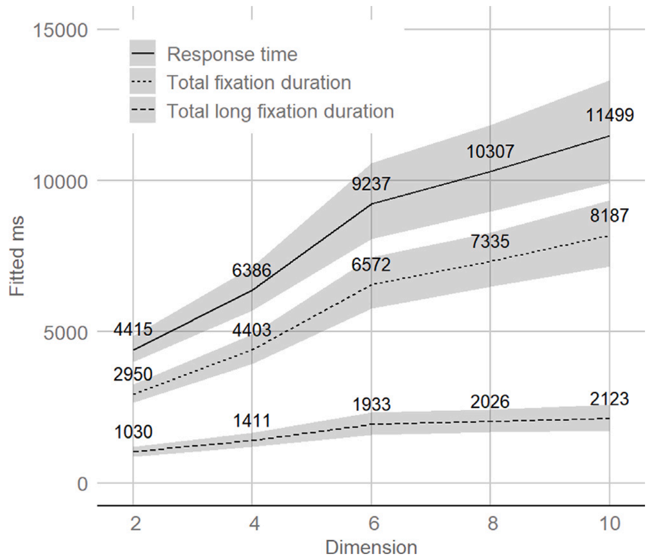


**Fig. 5.** Fitted values for RT, indicating effort (top); total duration of fixations (middle), indicating the sum of information intake and cognitive work; and total duration of long fixations (bottom), indicating cognitive work without information intake; in relation to the dimension of the bundles in the choice setting. The line from 2 to 6 dimensions appears slightly nonlinear due to the Box–Cox transformation necessitated by the model assumptions (please see the Supplementary material for details).

different products was more difficult'' (questions C10–C12 in Table A.3) when the number of product types was high vs. medium vs. low (M = 5.7 vs. 3.9 vs. 1.8; $p < 0.001$ for all pairwise comparisons). In other words, the participants consciously experienced more burden when the number of product types was higher. Very few participants (4/50) disagreed that a higher number of product types made the task more difficult, or that a lower number of products made it easier (2/50), or that the task was most difficult when the bundles were equally attractive (4/50).

In summary, we conclude that the participants did experience more cognitive burden (CL and the associated negative feeling) when the dimension of the bundles in the choice setting was higher. This supports hypothesis H1 (''Cognitive burden increases with the number of objectives'').

### 5.2. How did behavior change with higher dimension?

Fig. 5 shows the fitted models for RT and two eye-tracking measures (total fixation duration and total long fixation duration) as functions of dimension. The absolute values of these measures depend on the task and the individual, whereby we are only interested in changes in these measures when dimension is increased. For each of the three measures, the piecewise model with a bend at a dimension level of 8 had the best fit.

RT (top curve in Fig. 5) is a direct measure of effort spent on the task — the more work the participant puts in, the more time it takes. Here, the piecewise model has a better fit than a linear one, indicating that after a particular threshold (6 product types), the increase of time spent on the choice task is less steep than before it. This implies that while cognitive burden still increased the time spent on the task, the participants began making relatively quicker decisions, with less effort spent per piece of information, when the bundles had a high number of product types. On average, a trial with two product types in each bundle lasted about 4.5 s, while a trial with ten product types in each bundle lasted almost 12 s. The model also reveals that participants generally made decisions much quicker towards the end of the experiment; the difference between RT in the first and the last trial was on average 4.6 s, suggesting a notable learning effect.[9]

The two eye-tracking duration measures demonstrate the behavior change in more detail. When the trial time (RT) is directly determined by how much processing the task takes, the total duration of fixations is directly proportional to it. In our dataset, RT was heavily correlated with the total duration ($r$ = 0.964), indicating that when RT increases with the number of product types in each bundle, the gaze must visit more locations to gain enough information to make the choice. This provides support to our conclusion about RT. Using total duration of fixations (middle curve in Fig. 5), we find that the form of the graph closely replicates that of RT, including the bend that indicates the strategy change, and that the change between dimensions is proportionate between the two models. The difference between the top and middle curves represents the duration of saccades and blinks.

The bottom curve in Fig. 5 represents the total duration of longer fixations, which are indicative of effort focused on deep processing, without the search behavior. Notably, the curve shows the same pattern as the curves for RT and total fixation duration, but the right side of the graph is practically flat, indicating that next to none extra deep processing was done in choice tasks with ten product types compared to six or eight, despite the notable increase in information. The difference between the bottom curve and the two top curves suggests that the extra effort spent on the higher levels of dimensions (according to RT and total fixation duration) does not involve deeper processing. In other words, while more time is spent when the decision-maker must consider more information on the highest dimension level, it is spent on looking for relevant information, instead of deep processing the information in order to make a better decision.

This is corroborated by the post-experiment questionnaires where the participants reported their own views about their behavior using a Likert scale 1–7 (1 = Strongly disagree, 7 = Strongly agree; see Table A.3 in Appendix A). Participants agreed with the statement ''I decreased my effort in order to feel more comfortable'' (questions C13–C15 in Table A.3) when the number of product types was high vs. medium vs. low, i.e., they responded to burden by switching the choice strategy to something less burdensome[10] (M = 4.3 vs. 3.2 vs. 2.3, $p$ = 0.001 and $p$ = 0.009 for the pairwise comparisons between subsequent levels). Very few participants disagreed that they managed the difficulty by changing goals or strategy (6/48), or that they made up a decision rule to simplify the task (5/48).

Was the change in their behavior conscious? They agreed with ''I only considered some of the products and ignored the rest'' (questions C19–C21 in Table A.3) more when the number of product types was high vs. medium vs. low (M = 5.5 vs. 3.9 vs. 2.3, $p < 0.001$ for each pairwise comparison). That is, the participants agreed that they lessened the burden by ignoring some products altogether. They agreed with ''I ignored the number of copies for products'' (questions C22–C24 in Table A.3) more when the number of products was high vs. medium vs. low (M = 3.6 vs. 2.7 vs. 2.0; $p$ = 0.008 for the first pairwise comparison and $p$ = 0.09 for the latter, implying that the

[9] A possible explanation to this effect could be that our subjects were getting tired. Yet, this does not seem to have been the case: the average response to question B7 (''During the task to what extent did you feel tired?'') on a scale 1–9 (where 1 = Not at all and 9 = Very much) increased from 3.10 to 4.04 between the first and last block, a minor difference compared to the standard deviation of the responses which was ca. 2.0 in each block.

[10] Interestingly, the responses to the statement ''I tried to make each decision as carefully as possible'', filled in during the experiment (every 12 trials; question B15 in Table A.2 in Appendix A), did not corroborate this. There was a very small decrease in agreement from 2 to 6 dimensions (0.3 on scale from 1 to 9), and no difference at all between dimensions 6, 8, and 10. It is possible that ''as carefully as possible'' is too vague a statement and did not capture the switch in strategy, or that the participants considered the strategy change to be what was ''possible''.

difference between responses for medium and low numbers of products was statistically insignificant). In other words, participants lessened the burden by ignoring how many copies of each product type there were in the bundle and instead only considered which product types were involved. However, ignoring some product types was more prevalent than ignoring the number of copies.

To the question “If all products did not equally contribute to your choice, about how many products were the most influential?” (questions C25–C27 in Table A.3) the participants responded, on average, with M = 3.6 when the number of product types was high versus M = 2.6 when the number of product types was medium ($p < 0.001$). For a large majority of the participants, the number of product types considered was at most four. Even when the number of product types presented was high, only seven participants (14.5%) reported that they considered five ($n = 6$) or six ($n = 1$) types. No participant reported that more than 6 product types contributed to the decision.

The participants also reported that they continuously adapted their goals and strategies rather than changed their strategy at a certain point ($n = 35$ agreed versus $n = 8$ disagreed). Some participants also answered that “I made my decisions essentially randomly” (questions C16–C18 in Table A.3) more when the number of products was high vs. medium vs. low (M = 2.6 versus 1.9 vs. 1.4; $p = 0.007$ for the first pairwise comparison and $p = 0.22$ for the latter). That is, some participants lessened the burden by simply giving up. Ten participants agreed with this statement when the number of product types was high.

In conclusion, we found that in response to a higher dimension of the bundles in the choice setting, the participants changed their behavior by increasingly using a strategy in which they only considered a limited number of product types and ignored the rest, even as that (ostensibly) meant getting an inferior task reward. This is supported by how the participants reported they acted, but also by the objective physiological eye-tracking measures. These findings support hypothesis H2 (“When the number of objectives is high, decision-makers adapt to anticipated negative affect by considering a limited number of objectives only”).

### *5.3. How did higher dimension affect decision quality?*

Decision quality was measured through duplicate and preference consistency. Fig. 6 illustrates duplicate consistency – the percentage of identical choices in duplicate trials – as a function of dimension. Based on this figure, duplicate consistency was found to decrease linearly by ca. 10%-points when the dimension of bundles in the choice setting changed from 2 to 10. This suggests that the psychological burden increasing with dimension made it more difficult to make consistent choices. Yet, contrary to our results on RT and eye-tracking measures (see Fig. 5), a piecewise linear model did not fare any better for the duplicate consistency measure than a linear one. This suggests that the probability of making inconsistent choices increases with more information regardless of whether that information is used in decision-making — likely because it still influences the search process. However, with block-level data the amount of data points was an order of magnitude smaller than in the analyses on RT or eye-tracking measures, whereby this conclusion is more uncertain.

Preference consistency was measured as the share of choice settings in which the participants chose the bundle that maximized their proxy value function $V_i(\boldsymbol{x}) = \sum_{j=1}^{J} w_{ij} x_j$. Regression models for preference consistency were estimated for the case in which the weights $w_{ij}$ were equal to the participants' self-reported ratings for different product types, and for three sets of surrogate weights computed based on these ratings (rank sum weights, rank reciprocal weights, and rank-order centroid weights; cf. (6)–(8)). Among the surrogate weights, rank reciprocal weights resulted in the best model fit and also the highest average level of preference consistency (see the Supplementary material for details). Thus, Fig. 7 shows preference consistency – as a function of the dimension of the bundles in the choice setting – only with $w_{ij} = w_{ij}^{RR}$ (rank reciprocal weights, solid line) and $w_{ij} = r_{ij}$ (self-reported ratings, dotted line). In the case of self-reported ratings, preference consistency remains essentially constant (at around 68%) until a dimension of 8, after which it dips. For rank reciprocal weights, preference consistency increases from two to four dimensions, after which it starts to decrease, indicating that best consistency is reached at a dimension level of four. It also appears that rank reciprocal weights capture the subjects' choices a bit more accurately than the self-reported ratings as such in that rank reciprocal weights yield higher preference consistency values at all dimension levels.

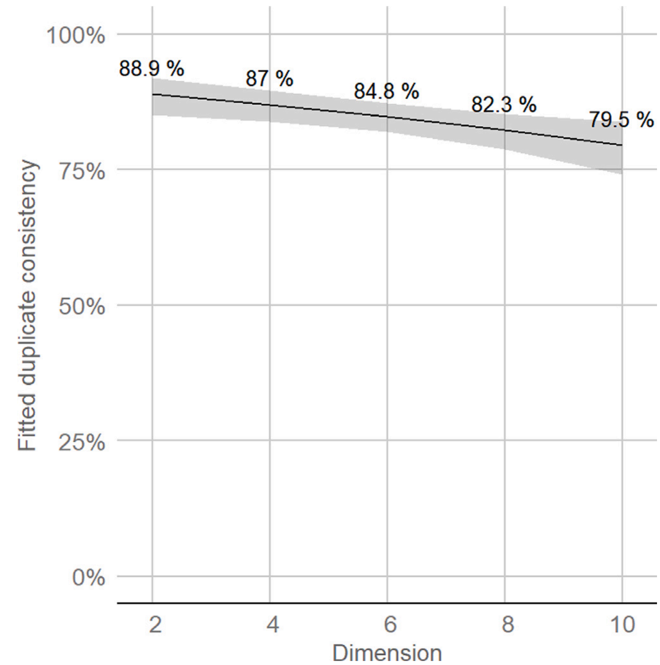


**Fig. 6.** Duplicate consistency in relation to the dimension of the bundles in the choice setting.

Despite differences in detail, both curves in Fig. 7 suggest that making choices that are consistent with one's preferences becomes more difficult when the number of dimensions increases. We interpret that this is due to the participants making choices based on only a few product types. Fasolo et al. (2007) have shown that when the objective-specific performances of decision alternatives are negatively correlated, the consideration of only a few objectives in decision making leads to poor choices. Such negative correlations are necessary when only non-dominated alternatives are considered, which is the case in our experimental study. Hence, when the decision is made according to only a few product types out of many, the ignored portion of the value of the bundle becomes large enough to decrease preference consistency.

It should be emphasized that our results on preference consistency depend on rather strong assumptions: (i) that the participants are able to reasonably accurately rate their preferences for different product types using a discrete scale between 1–9 and (ii) that the overall value of a bundle is captured by the linear-additive proxy value function (5) (with subject-specific weights for different product types corresponding either to the self-reported ratings for these product types or to surrogate weights computed based on these ratings). Regarding the first assumption, on average the participants reported that their preferences were clear to them (mean response M = 5.8 on the scale from 1 = Strongly disagree to 7 = Strongly agree, with only 4 participants disagreeing; see question C4 in Table A.3 in Appendix A) — in other words, they reported that they knew what they liked. Yet, they considered some products to have negative value (M = 5.0, with only 8 participants disagreeing; see question C6 in Table A.3), which

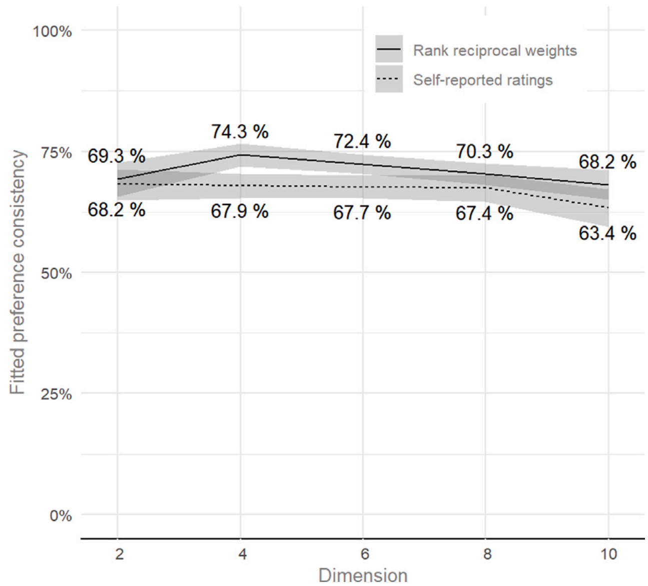


**Fig. 7.** Predicted preference consistency in relation the dimension of the bundles in the choice setting with $w_{ij} = w_{ij}^{RR}$ (rank reciprocal weights, solid line) and $w_{ij} = r_{ij}$ (self-reported ratings, dotted line).

was something we anticipated but did not take into account when designing the experiment. The second assumption is unlikely to hold for each individual participant, whereby the numerical results of Fig. 7 should be approached with caution. However, the qualitative result of preference consistency dipping after the dimension of the bundles in the choice setting becomes high enough is aligned with earlier literature and may, therefore, be held as valid.

In summary, decision consistency generally decreases when the dimension of the bundles in the choice setting increases. Furthermore, when the dimension is high enough, the ability to choose the alternative that would be aligned with one's preferences (at least when these preferences are assumed to be represented by the proxy value function) becomes notably more difficult. These findings support our hypothesis H3 ("Decision consistency decreases with the number of objectives").

### *5.4. Did a specific dimension level exist beyond which it was too burdensome to make high-quality decisions?*

Many of our results (namely on blink rate, RT, eye-tracking measures, and preference consistency when using rank-reciprocal weights computed from the participants' self-reported ratings for different product types in the proxy value functions) suggested that there was a bend in the participants' behavior and decision consistency at a dimension level of four or six, indicating that decision-making could have become too burdensome at this point. Yet, it should be noted that individual differences regarding the specific location of the bend were large, as demonstrated by the relatively wide confidence intervals and the fact that random slopes models largely had a better fit than random intercepts only models. Finally, it is likely that the specific location of the bend is highly context-specific and could have been different for different choices of product types, bundles, and individual participants. In summary, no strong conclusion based on this study can be made about a specific dimension level after which high-quality decisions can no longer be made — only that there is a limit, and it is often quite low.

## 6. Conclusions

The aim of this paper has been to investigate the impact of the number of objectives on the cognitive burden experienced by human decision-makers in multiobjective decision settings, and the effects that the cognitive burden has on behavior and decision consistency. For this purpose, we have carried out an experimental laboratory study in which human subjects have made pairwise comparisons between product bundles of varying dimension. In this experimental study, we have used psychophysiological, behavioral, and self-report methods. Our results have shown that increasing the dimension of the bundles in the choice setting increases cognitive load, which is a source of discomfort that decision-makers seek to avoid. In response, there is a tendency for humans to simplify the more complex comparisons, especially by ignoring all but a selected few of the dimensions. This tendency is more likely when the number of dimensions is high. The increase in cognitive burden goes hand in hand with somewhat decreasing consistency of decision-making, which can be seen as an indication of lower decision quality.

Because the human body does not have definite markers for internal processes such as cognitive burden or decision-making strategies, all the methods have measured different facets about cognitive burden and its impacts on behavior, and all the measures have their own sources of noise. The actual behavioral processes depend on the individual and on some inherent properties of the measures (e.g., physical limitations of eye movements), and are subject to some random variation. In addition, the details of the decisions themselves (e.g., subjective or objective criteria, complex vs. simple information, pairwise choice or one out of many) and the decision-making situation (e.g., time limit or not, external pressure) are likely to influence the results. Hence, the specific numerical results of our study (e.g., a bend in the piecewise linear models at on average 4–6 dimensions indicating a significant change in behavior after this point) should not be taken as universal rules. Rather, the takeaway of our study is the general tendency of the participants to focus on only a few selected objectives (i.e., product types) at perhaps surprisingly low dimension levels, and to simply ignore a large portion of the available information.

We believe that the phenomenon of simplifying a complex decision task is a result from a process in which the optimization of task-unrelated goals (such as comfort) leads to ceasing to optimize the task-related goals and adopting less cognitively demanding strategies that still produce decisions that are good enough. Nevertheless, it should be noted that the difficulty of a decision-making task is not an objective phenomenon, but rather a result of interaction between the psychological capacities, motivations, and characteristics of the participant, and the requirements of the task. This means that the reactions and resulting behavior patterns may differ significantly between individuals (see, e.g., difference between maximizers and satisficers; Schwartz et al. 2002). However, as a physiological study in a high-cost laboratory environment, our sample size is not large enough to allow distinct groups according to individual differences.

Furthermore, it should be noted that the incentives to perform optimally in our experiment were not very strong: for most of us, even if we prefer a movie ticket to a can of cola, the decision is not so important that we would be willing to put a lot of effort into ensuring that we get the former and not the latter. We acknowledge that the number of objectives that decision-makers are likely to truly consider also depends on the incentives. Regardless, the limitations of human cognitive capabilities and the tendency of the human brain to minimize effort wherever possible do not disappear. Thus, although it may be possible to increase the number of truly considered objectives, it cannot be increased indefinitely, and not without decreasing decision quality, the decision-maker's feeling of comfort (and therefore long-term motivation), or both. Further empirical studies on this subject are required.

We also found that the effects in our experiment were ultimately relatively small, although statistically significant to establish the concluded pattern. In future studies, strengthening the incentives and increasing the number of data points (either by more repetitions or more participants) should make the differences more pronounced. Our experiment was already rather long and tiring for the participants, including tedious set-up, briefings, responding to surveys, and making 120 binary choices (trials) between product bundles. Moreover, the participants had to sit still during the experiment not to corrupt the eye movement data. Another option would have been to split the experiment into two parts. Both methods have their pros and cons and should be considered in future studies.

Another possible weakness in our study was that the cognitive and affective consequences of cognitive burden were measured only through self-reports on perceived complexity, information overload, and negative affect. We do not believe that relying on self-reports in this one aspect has influenced our conclusions, which are also supported by both theory and non-voluntary measures of blink rate, response time, and eye-tracking data. Yet, to further strengthen the reliability of the results, more cross-validation should be considered in future studies. Another interesting direction for future research is to carry out an experiment in which the subjects choose one out of many Pareto-optimal alternatives instead of only two. Specifically, the interaction between the number of alternatives and objectives in relation to cognitive burden is a topic that merits further exploration. Finally, efforts should be taken to study the impact of increasing the number of objectives on decision quality more thoroughly. In this paper we have focused on decision consistency, which does not fully capture the quality of decisions. The use of measures such as the proportion of value lost or the share of choice settings in which the most preferred alternative was chosen would require the elicitation of each subject's individual value function, which can be difficult and time-consuming. Another approach would be to include dominated alternatives in the choice settings and measure decision quality as the share of settings in which the dominant alternative was chosen.

Our study has significant implications for MCDM, as well as EMO studies. We conclude that, due to cognitive limitations of humans, it may not be worthwhile to unnecessarily focus on "too many" objectives in addressing many-objective optimization problem solving tasks. Introducing excessively many objectives involves a significant risk that decision-makers simply ignore large parts of information related to the decision, thereby defeating the purpose of using many objectives in the first place. A better strategy may be to focus on the most important (e.g., two to six) objectives, and perhaps include the remaining less important objectives via constraints into the mathematical formulation or use them as second-level objectives in a hierarchical structure (Lai, Fiaschi, Cococcioni, & Deb, 2021).

Our study also contributes to research on the use of heuristics in multiobjective optimization and decision-making tasks. Research efforts in this field have thus far been directed towards the use of simulation to investigate the conditions under which certain heuristics (such as focusing only on the most important objectives) perform well in terms of decision quality (e.g., Barron 1987, Barron and Barrett 1999, Barron and Kleinmuntz 1986, Fasolo et al. 2007). Our results complement this literature by providing experimental evidence pointing to a behavior like the use of heuristics under certain conditions. In this respect, the use of psychophysiological measurements is novel. Their distinct benefit is that the results are not dependent on what the participants consciously understand and admit doing, as the behavior can be inferred from their actions alone.

The conclusions of our paper are not dramatically different from what Miller (1956) wrote nearly 70 years ago. Humans face limits to their cognitive capabilities. Miller's Law predicts that the average person can only keep 7 (± 2) items in their working memory, which compares well to the pieces of information our participants willingly processed (note that with a dimension level of four, the participant needed to keep in memory the information about two bundles containing up to eight product types in total). More recently, Cowan (2010, 2012) has extensively researched the limits of working memory for all ages and concluded that working memory store for young adults is limited to 3–5 meaningful items. The purpose of the current study is not to define an exact limit on the number of dimensions that humans can deal with (because any results on such limits invariably depend on the specific ways in which experiments have been conducted) but, rather, to conclude that such a relatively low limit exists and influences the decision process and its results.

Decision-making literature generally talks about the decision-maker in abstract. The role of an individual and individual differences is under-researched, even ignored in the literature. As a part of future work, we intend to use the individual-level data, which was collected as part of our study, for investigating such individual differences. Other extensions include investigating (i) the effect of dimension in progressive (interactive) use of decision-making within MCDM and EMO algorithms, (ii) the impact of clustering many objectives into smaller groups and the use of each cluster in different rounds of decision-making sessions systematically, leading to more involved algorithmic changes of optimization methods, and (iii) the effect of dimension in group decision-making to reduce an individual decision-maker's psychological burden by sharing it among the group members. Finally, it would be interesting to study what kinds of strategies human decision-makers use in multiobjective decision settings, and how the use of these strategies depends on the number of objectives and decision-makers' individual traits. The current study paves the way for such additional studies, which hopefully would provide a more comprehensive understanding of the effect of objective dimensionality for multiobjective optimization and decision-making.

## CRediT authorship contribution statement

**J. Matias Kivikangas:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Eeva Vilkkumaa:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation. **Julian Blank:** Software, Methodology. **Ville Harjunen:** Investigation. **Pekka Malo:** Writing – review & editing, Validation, Software, Resources, Methodology, Funding acquisition, Formal analysis, Data curation. **Kalyanmoy Deb:** Writing – original draft, Supervision, Conceptualization. **Niklas J. Ravaja:** Methodology, Conceptualization. **Jyrki Wallenius:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

## Acknowledgments

We wish to express our thanks to Wihuri Foundation, HSE Support Foundation, and Professors Markku Kuula and Matti Rossi, Department of Information and Service Management, Aalto University School of Business, Finland, for financial support. We also wish to extend thanks to Ms. Zania Sovijärvi-Spape, our laboratory assistant, for her help in conducting the experiments. We also wish to thank Prof. Konstantinos Katsikopoulos (University of Southampton) and Giovanni Misitano (University of Jyväskylä) for their useful comments.

## Appendix A. Questionnaires used in the experiment

See Tables A.1–A.3.

## Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ejor.2024.10.039.

**Table A.1**
Questions for self-report measures before starting the experiment. The response scale for each question is 1–9 with interpretations given for the extreme values only.

| Label | Question | Interpretations for the extreme values of the response scale |
|---|---|---|
| A1 | How tired are you right now? | 1: Not at all<br>9: Extremely |
| A2 | How distracted are you right now by other things outside this experiment? | 1: Not at all<br>9: Extremely |
| A3 | How nervous are you right now? | 1: Not at all<br>9: Extremely |

**Table A.2**
Questions for self-report measures after each block. The response scale for each question is 1-9 with interpretations given for the extreme values only. For the pictorial SAM scales, please see Bradley and Lang (1994).

| Label | Question | Interpretations for the extreme values of the response scale |
|---|---|---|
| Considering the past twelve decisions you made... | | |
| B1 | How much effort did the task require? | 1: No effort at all<br>9: Extreme effort |
| B2 | How difficult was the task? | 1: Extremely easy<br>9: Extremely difficult |
| During the task, to what extent did you feel... | | |
| B3 | Attentive? | 1: Not at all<br>9: Very much |
| B4 | Uncomfortable? | 1: Not at all<br>9: Very much |
| B5 | Frustrated? | 1: Not at all<br>9: Very much |
| B6 | Bored? | 1: Not at all<br>9: Very much |
| B7 | Tired? | 1: Not at all<br>9: Very much |
| B8 | [Pictorial SAM valence scale]? | 1: Unhappy, unsatisfied<br>9: Happy, satisfied |
| B9 | [Pictorial SAM arousal scale]? | 1: Calm, drowsy, still<br>9: Energetic, agitated, alert |
| B10 | [Pictorial SAM dominance scale]? | 1: Being at the mercy of the situation, unable to control the situation<br>9: Being completely in control of the situation, able to control the situation |
| Considering the past twelve decisions you made, to what extent do you agree with the following statement? | | |
| B11 | The task seemed very complex | 1: Strongly disagree<br>9: Strongly agree |
| B12 | I was overwhelmed | 1: Strongly disagree<br>9: Strongly agree |
| B13 | I was motivated | 1: Strongly disagree<br>9: Strongly agree |
| B14 | I felt like not wanting to do the task | 1: Strongly disagree<br>9: Strongly agree |
| B15 | I tried to make each decision as carefully as possible | 1: Strongly disagree<br>9: Strongly agree |

**Table A.3**
Questions for post-experiment self-report measures. Subjects were not given definitions on what was meant by a low, medium, or high number of products. The response scale for questions C1–C24 is 1–7 with interpretations given for the extreme values only.

| Label | Question | Interpretations for the extreme values of the response scale |
|---|---|---|
| C1 | The task was easy when there were only few product types in the bundles | 1: Strongly disagree<br>7: Strongly agree |
| C2 | The task was difficult when there were many product types in the bundles | 1: Strongly disagree<br>7: Strongly agree |

(continued on next page)

**Table A.3** (*continued*).

| Label | Question | Interpretations for the extreme values of the response scale |
|---|---|---|
| C3 | I managed the task difficulty by changing my goals or strategy | 1: Strongly disagree<br>7: Strongly agree |
| C4 | I had clear preferences between the product types | 1: Strongly disagree<br>7: Strongly agree |
| C5 | The decision was most difficult when both bundles had equally attractive (or unattractive) products | 1: Strongly disagree<br>7: Strongly agree |
| C6 | Some product types I actively didn't want; when they were in the bundle, it decreased the bundle's value for me | 1: Strongly disagree<br>7: Strongly agree |
| C7 | I made up a decision rule in my mind to simplify the task | 1: Strongly disagree<br>7: Strongly agree |
| C8 | Please indicate which statement is closer to your experience about the task | 1: The task became more difficult towards the end<br>7: The task became more easy towards the end |
| C9 | Please indicate which statement is closer to your experience about the task | 1: I experienced a clear point where I changed my goals or strategy<br>7: I continuously adapted my goals or strategies to each decision |
| Looking at and keeping in mind all the different products was difficult | | |
| C10 | When there was a high number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| C11 | When there was a medium number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| C12 | When there was a low number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| I decreased my effort in order to feel more comfortable | | |
| C13 | When there was a high number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| C14 | When there was a medium number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| C15 | When there was a low number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| I made my decision essentially randomly | | |
| C16 | When there was a high number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| C17 | When there was a medium number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| C18 | When there was a low number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| I only considered some of the products and ignored the rest | | |
| C19 | When there was a high number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| C20 | When there was a medium number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| C21 | When there was a low number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| I ignored the number of copies for products | | |
| C22 | When there was a high number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| C23 | When there was a medium number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| C24 | When there was a low number of product types in the bundle | 1: Strongly disagree<br>7: Strongly agree |
| If all product types did not equally contribute to your decision, about how many product types were the most influential | | |
| C25 | When there was a high number of product types in the bundle? | [Insert number] |
| C26 | When there was a medium number of product types in the bundle? | [Insert number] |
| C27 | When there was a low number of product types in the bundle? | [Insert number] |

## References


Afsar, B., Silvennoinen, J., Misitano, G., Ruiz, F., Ruiz, A. B., & Miettinen, K. (2022). Designing empirical experiments to compare interactive multiobjective optimization methods. *Journal of the Operational Research Society*, *74*, 1–12.

Agarwal, M. K., & Chatterjee, S. (2003). Complexity, uniqueness, and similarity in between-bundle choice. *Journal of Product & Brand Management*, *12*(6), 358–376.

Arkes, H. R., Gigerenzer, G., & Hertwig, R. (2016). How bad is incoherence? *Decision*, *3*(1), 20–39.

Arrow, K. J. (1982). Risk perception in psychology and economics. *Economic Inquiry*, *20*(1), 1.

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23.

Barrett, L. F., & Finlay, B. L. (2018). Concepts, goals and the control of survival-related behaviors. *Current Opinion in Behavioral Sciences*, *24*, 172–179.

Barron, F. H. (1987). Influence of missing attributes on selecting a best multiattributed alternative. *Decision Sciences*, *18*(2), 194–205.

Barron, F. H., & Barrett, B. E. (1999). Linear inequalities and the analysis of multi-attribute value matrices. In *Decision science and technology: Reflections on the contributions of ward edwards* (pp. 211–225). Springer.

Barron, F. H., & Kleinmuntz, D. (1986). Sensitivity in value loss of linear multiattribute value models to attribute completeness. In *New directions: Research in decision making*. Elsevier.

Baucells, M., Carrasco, J. A., & Hogarth, R. M. (2008). Cumulative dominance and heuristic performance in binary multiattribute choice. *Operations Research*, *56*(5), 1289–1304.

Bawden, D., & Robinson, L. (2009). The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, *35*(2), 180–191.

Bellemare, C., Bissonnette, L., & Kröger, S. (2014). Statistical power of within and between-subjects designs in economic experiments.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *26*(2), 211–243.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59.

Browne, G. J., Pitts, M. G., & Wetherbe, J. C. (2007). Cognitive stopping rules for terminating information search in online tasks. *MIS Quarterly*, *31*, 89–104.

Cacioppo, J. T., Tassinary, L. G., Berntson, G. G., et al. (2007). Psychophysiological science: Interdisciplinary approaches to classic questions about the mind. *vol. 3*, In *Handbook of psychophysiology* (pp. 1–16).

Cardona, G., & Quevedo, N. (2014). Blinking and driving: the influence of saccades and cognitive workload. *Current Eye Research*, *39*(3), 239–244.

Chikhi, S., Matton, N., & Blanchet, S. (2022). EEG power spectral measures of cognitive workload: A meta-analysis. *Psychophysiology*, *59*(6), Article e14009.

Coello, C. A. C. (2007). *Evolutionary algorithms for solving multi-objective problems*. Springer.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57.

Cowan, N. (2012). *Working memory capacity*. Psychology Press.

Danielson, M., & Ekenberg, L. (2017). A robustness study of state-of-the-art surrogate weights for MCDM. *Group Decision and Negotiation*, *26*, 677–691.

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. Wiley.

Deb, K., & Jain, H. (2013). An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, *18*(4), 577–601.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*(2), 182–197.

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin Books.

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, *100*(1), 223–234.

Dyer, J. S., & Sarin, R. K. (1979). Measurable multiattribute value functions. *Operations Research*, *27*(4), 810–822.

Eder, A. B., Elliot, A. J., & Harmon-Jones, E. (2013). Approach and avoidance motivation: Issues and advances. *Emotion Review*, *5*(3), 227–229.

Fasolo, B., McClelland, G. H., & Todd, P. M. (2007). Escaping the tyranny of choice: When fewer attributes make choice easier. *Marketing Theory*, *7*(1), 13–26.

Feldon, D. F., Callan, G., Juth, S., & Jeong, S. (2019). Cognitive load as motivational cost. *Educational Psychology Review*, *31*(2), 319–337.

Figueira, J. R., Greco, S., Mousseau, V., & Słowiński, R. (2008). Interactive multi-objective optimization using a set of additive value functions. In *Multiobjective optimization: Interactive and evolutionary approaches* (pp. 97–119). Springer.

Fonseca, C. M., & Fleming, P. J. (1993). Genetic algorithms for multiobjective optimization: formulation, discussion and generalization. *vol. 93*, In *Genetic algorithms: Proceedings of the fifth international conference (s. forrest, ed.), san mateo, CA* (pp. 416–423).

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. USA: Oxford Univ Press.

Glöckner, A., & Herbold, A.-K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, *24*(1), 71–98.

Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. Oxford University Press.

Hammond, K. R. (2007). *Beyond rationality: The search for wisdom in a troubled time*. Oxford University Press.

Harrington, N. (2005). The frustration discomfort scale: development and psychometric properties. *Clinical Psychology & Psychotherapy*, *12*(5), 374–387.

Huband, S., Barone, L., While, L., & Hingston, P. (2005). A scalable multi-objective test problem toolkit. In *Evolutionary multi-criterion optimization: Third international conference, EMO 2005, guanajuato, Mexico. proceedings 3*. Springer.

Iyengar, R., Ansari, A., & Gupta, S. (2007). A model of consumer learning for service quality and usage. *Journal of Marketing Research*, *44*(4), 529–544.

Kadziński, M., Greco, S., & Słowiński, R. (2012). Selection of a representative value function in robust multiple criteria ranking and choice. *European Journal of Operational Research*, *217*(3), 541–553.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Katsikopoulos, K. V., Durbach, I. N., & Stewart, T. J. (2018). When should we use simple decision models? A synthesis of various research strands. *Omega*, *81*, 17–25.

Köksalan, M. M., Wallenius, J., & Zionts, S. (2011). *Multiple criteria decision making: From early history to the 21st century*. World Scientific.

Korhonen, P. J., & Wallenius, J. (2020). *Making better decisions*. Springer.

Korhonen, P. J., Wallenius, J., Genc, T., & Xu, P. (2021). On rational behavior in multi-attribute riskless choice. *European Journal of Operational Research*, *288*(1), 331–342.

Korhonen, P., Wallenius, J., & Zionts, S. (1984). Solving the discrete multiple criteria problem using convex cones. *Management Science*, *30*(11), 1336–1345.

Kourouxous, T., & Bauer, T. (2019). Violations of dominance in decision-making. *Business Research*, *12*, 209–239.

Krantz, D., Luce, D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement, vol. I: Additive and polynomial representations*. New York: Academic Press.

Lai, L., Fiaschi, L., Cococcioni, M., & Deb, K. (2021). Solving mixed pareto-lexicographic multiobjective optimization problems: the case of priority levels. *IEEE Transactions on Evolutionary Computation*, *25*(5), 971–985.

Li, X., Luh, D.-B., & Chen, Z. (2024). A systematic review and meta-analysis of eye-tracking studies for consumers' visual attention in online shopping. *Information Technology and Control*, *53*(1), 187–205.

Linkov, I., Moberg, E., Trump, B. D., Yatsalo, B., & Keisler, J. M. (2020). *Multi-criteria decision analysis: Case studies in engineering and the environment*. CRC Press.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical surveys*. Wiley New York.

Malhotra, N. K. (1982). Information load and consumer decision making. *Journal of Consumer Research*, *8*(4), 419–430.

March, J. G. (1991). How decisions happen in organizations. *Human-Computer Interaction*, *6*(2), 95–117.

March, J. G. (2006). Rationality, foolishness, and adaptive intelligence. *Strategic Management Journal*, *27*(3), 201–214.

Methling, F., Abdeen, S. J., & von Nitzsch, R. (2022). Heuristics in multi-criteria decision-making: the cost of fast and frugal decisions. *EURO Journal on Decision Processes*, *10*.

Miao, M., & Jayakar, K. (2014). Bounded rationality and consumer choice: an evaluation of consumer choice of mobile bundles. *Chinese Journal of Communication*, *7*(2), 191–211.

Miettinen, K. (1999). *Nonlinear multiobjective optimization*. Springer Science & Business Media.

Miettinen, K., Ruiz, F., & Wierzbicki, A. P. (2008). Introduction to multiobjective optimization: Interactive approaches. In *Multiobjective optimization: Interactive and evolutionary approaches* (pp. 27–57). Springer.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Negi, S., & Ritayan, M. (2020). Fixation duration and the learning process: An eye tracking study with subtitled videos. *Journal of Eye Movement Research*, *13*(6), 1–15.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2016). Cognitive load measurement as a means to advance cognitive load theory. In *Cognitive load theory* (pp. 63–71). Routledge.

Paas, F., Van Gog, T., & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review*, *22*, 115–121.

Pande, S. M., Papamichail, K. N., & Kawalek, P. (2021). Compatibility effects in the prescriptive application of psychological heuristics: Inhibition, integration and selection. *European Journal of Operational Research*, *295*(3), 982–995.

Peng, M., Xu, Z., & Huang, H. (2021). How does information overload affect consumers' online decision process? An event-related potentials study. *Frontiers in Neuroscience*, *15*, Article 695852.

Pennington, R. R., & Kelton, A. S. (2016). How much is enough? An investigation of nonprofessional investors information search and stopping rule use. *International Journal of Accounting Information Systems*, *21*, 47–62.

Perkhofer, L., & Lehner, O. (2019). Using gaze behavior to measure cognitive load. In *Information systems and neuroscience* (pp. 29–73).

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17–35.

Phelps, S., & Köksalan, M. (2003). An interactive evolutionary metaheuristic for multiobjective combinatorial optimization. *Management Science*, *49*(12), 1726–1738.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Roetzel, P. G. (2019). Information overload in the information age: a review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development. *Business Research*, *12*(2), 479–522.

Rutkowski, A.-F., & Saunders, C. S. (2010). Growing pains with information overload. *Computer*, *43*(06), 94–96.

Rutkowski, A.-F., & Saunders, C. (2018). *Emotional and cognitive overload: The dark side of information technology*. Routledge.

Savolainen, R. (2007). Filtering and withdrawing: strategies for coping with information overload in everyday contexts. *Journal of Information Science*, *33*(5), 611–621.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: happiness is a matter of choice. *Journal of Personality and Social Psychology*, *83*(5), 1178–1197.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*, 99–118.

Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N. S., et al. (2021). Review of eye tracking metrics involved in emotional and cognitive processes. *IEEE Reviews in Biomedical Engineering*, *16*, 260–277.

Srinivas, N., & Deb, K. (1994). Muiltiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, *2*(3), 221–248.

Struk, A. A., Carriere, J. S., Cheyne, J. A., & Danckert, J. (2016). A short boredom proneness scale. *Assessment*, *24*(3), 346–359.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 92–114.

Wallenius, J., Dyer, J. S., Fishburn, P. C., Steuer, R. E., Zionts, S., & Deb, K. (2008). Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. *Management Science*, *54*(7), 1336–1349.

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*: *Technical report*, University of Iowa.

Wilbanks, B. A., Aroke, E., & Dudding, K. M. (2021). Using eye tracking for measuring cognitive workload during clinical simulations: literature review and synthesis. *CIN: Computers, Informatics, Nursing*, *39*(9), 499–507.

Zhang, Q., & Li, H. (2007). MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, *11*(6), 712–731.

Zhang, Q., Zhou, A., Zhao, S., Suganthan, P. N., Liu, W., & Tiwari, S. (2008). *Multi-objective optimization test instances for the CEC 2009 special session and competition*: *Technical report*, University of Essex.

Zitzler, E. (1999). *Evolutionary algorithms for multiobjective optimization: Methods and applications*. Shaker Ithaca.