

# Does inducing choice procedures make individuals better off? An experimental study

Luigi Mittone<sup>1</sup> · Mauro Papi<sup>2</sup>

Published online: 30 January 2017

© The Author(s) 2017. This article is published with open access at Springerlink.com

**Abstract** Over the years the psychological and economic literature on multi-attribute individual decision-making has focused its attention on examining what class of heuristics better describes subjects' behaviour. In contrast, motivated by the proliferation of online choice platforms, we investigate whether inducing subjects to use holistic vs. characteristic-based search (CBS) procedures has an effect on the quality of their decision by proposing a between-subject experiment involving an innovative visual choice task. We find that encouraging subjects to use CBS heuristics as opposed to holistic ones makes them better off. We also examine how subjects' performance is related with complexity, time pressure, and random choice by running simulations and link our results to the related literature.

---

We would like to thank Marco Tecilla for excellent computer programming. We are extremely grateful to an anonymous referee for her/his support to our paper, the many insightful comments, and the valuable suggestions on the research program. We also thank Miguel Costa-Gomes, Paola Manzini, Ariel Rubinstein, the participants to the 2016 International Meeting on Experimental and Behavioral Social Sciences (Rome), the seminar audiences at the universities of Aberdeen, St Andrews and Trento and the editor of this journal. Financial support from the University of Trento's core funding is gratefully acknowledged. Any error is our own responsibility.

---

**Electronic supplementary material** The online version of this article (doi:[10.1007/s11238-017-9590-3](https://doi.org/10.1007/s11238-017-9590-3)) contains supplementary material, which is available to authorized users.

---

✉ Mauro Papi  
m.papi@abdn.ac.uk

Luigi Mittone  
luigi.mittone@unitn.it

<sup>1</sup> Cognitive and Experimental Economics Laboratory, University of Trento, via Verdi, 26, 38100 Trento, Italy

<sup>2</sup> School of Business, University of Aberdeen, Edward Wright Building, Dunbar Street, Aberdeen AB243QY, UK

**Keywords** Between-subject · Characteristic-based search · Heuristic · Holistic · Multi-attribute problem · Simulations

## 1 Introduction

Over the past years innovative online choice platforms have been developed to facilitate the consumer's purchasing decisions. For example, a consumer interested in purchasing a high-performance sports car can visit the Ferrari's web-site and use the car configurator, which allows the consumer to design his ideal car by shaping its characteristics, such as colour, seat type and, wheel cap.<sup>1</sup> Another example is given by web-sites that through dedicated web-search engines allow the consumer to shortlist the set of available products by specifying the characteristics that the desired product should possess. Rightmove.co.uk, for instance, is a British website specialized in flat renting and gives the consumer the opportunity to shortlist on the basis of attributes, such as location, number of bedrooms, and rental price.<sup>2</sup>

Such choice platforms induce the consumer to follow certain choice procedures. In the Ferrari's car configurator case, the consumer is asked to construct his most preferred product by combining a set of available attributes. In the Rightmove.co.uk's case, on the other hand, the consumer has to shortlist the set of available products by specifying what properties his most desirable product should satisfy. In general the nature of the choice procedure utilized by a decision-maker might have an effect on the outcome of the decision and, as a result, on the welfare of the decision-maker himself. So far the psychological and economic literature on multi-attribute individual decision-making has focused its attention on examining what choice procedure better describes the subjects' behaviour.<sup>3</sup> In contrast, motivated by the continuing expansion of online choice platforms, we investigate whether inducing subjects to use holistic vs. characteristic-based search (CBS) procedures makes them better off.

Holistic procedures are procedures according to which the decision-maker examines the attributes within alternatives. Utility maximisation is an example of a holistic procedure, as a rational decision-maker first examines the attributes within an alternative (e.g. the prizes and the corresponding probabilities of a lottery), 'attaches' a utility value to it (e.g. expected utility), and then examines the next alternative. Another example of a holistic procedure is the satisficing heuristic (Simon 1955). CBS procedures, on the other hand, are procedures according to which the decision-maker examines the attributes across alternatives (Payne et al. 1993). The lexicographic and elimination-by-aspects (Tversky 1972) procedures are CBS examples. A decision-maker following a CBS procedure focuses his attention on one dimension (or various dimensions) only and discards all alternatives that are dominated on that dimension or do not meet a certain pre-determined threshold. Unlike traditional supermarkets,

---

<sup>1</sup> See figure 1 in the supplementary material.

<sup>2</sup> See figure 2 in the supplementary material.

<sup>3</sup> See Gigerenzer et al. (2011) as an example in the psychological literature and Gabaix et al. (2006) as an example in the economic literature. We will extensively discuss the related psychological and economic literature in Sect. 6.

online choice platforms, such as the ones described above, typically induce procedures that encompass CBS elements.<sup>4</sup>

Our research question is relevant for both psychologists and economists for various reasons. First, understanding whether inducing individuals to use different cognitive-comparative processes of the alternative's characteristics has an effect on the quality of their decision is a key problem. Second, it is interesting to examine whether or not subjects perform better by using a class of heuristics (i.e., CBS) that are generally inconsistent with the application of the utility maximisation procedure. Third, given the growing interest within economic theory about rational and boundedly rational choice procedures, our experiment 'searches for facts' within the domain of multi-attribute decision problems.<sup>5</sup>

In this paper we propose a between-subject design in which subjects are asked to perform the same choice task by inducing them to use different procedures. Our experiment consists of an innovative visual choice task, whereby subjects are shown a target alternative—an abstract figure—and asked to select or construct the alternative that most closely looks like the target. The baseline treatment induces a holistic procedure by asking subjects to *select* the figure that most closely looks like the target among those available. In contrast, two other treatments, which we call *build* and *destroy*, induce—in different ways—subjects to use characteristic-based search (CBS) procedures by asking them to *construct* the figure the most closely looks like the target figure by combining the blocs available. Across all treatments we vary both the time pressure level—interpretable as the search cost—and the complexity of the choice task.

It is worth emphasizing that in this experiment we fully control subjects' preferences, which we induce via monetary incentives. This is not because we rule out the possibility that inducing individuals to use certain choice procedures does not have an effect on their preferences. On the contrary, we believe that the preference-formation issue is relevant within the broader context of our research question. However, we think that our methodology is appropriate for at least two reasons. First, given that our aim is to understand whether inducing choice procedures has an effect on individuals' welfare, we need to know their preferences to be able to make welfare judgments. Inducing preferences via monetary incentives is a standard technique in experimental economics to achieve this goal (Smith 1976; Camerer and Hogarth 1999). Second,

---

<sup>4</sup> The Rightmove.co.uk's web-search engine induces an heuristic that resembles a CBS, because it implements a shortlisting procedure in the sense that it encourages the consumer to think about characteristics that he likes and discards all flats that do not possess those, such as a rental price greater than £850 per calendar month, less than two bedrooms, and a location more than three miles away from the desired one. The Ferrari's car configurator also induces a heuristic that encompasses CBS elements, as it induces the consumer to focus his attention on the car's characteristics sequentially (e.g. the choice of the colour first, then that of the wheel cap, then that of the seat and so on) with the objective of creating one alternative. On the contrary, in traditional supermarkets the available products are already constructed and—leaving aside some generic labels on the shelves to direct attention—consumers are not induced to use shortlisting procedures.

<sup>5</sup> This theoretic literature includes, for instance, Manzini and Mariotti (2007) and Apesteguia and Ballester (2013) that propose a model in which the decision-maker uses a shortlisting procedure. Rubinstein and Salant (2006) and Caplin and Dean (2011), on the other hand, formalize the satisficing heuristic proposed by Simon (1955).

since our study is—to the best of our knowledge—the first to investigate the effects of inducing individuals to use certain choice procedures, it is natural to start with an experimental design over which we have as much control as possible to be able to isolate the different effects.<sup>6</sup> The natural following step of this research, which we are already working on, is to extend the current experimental design with the objective of examining whether (and possibly how) preferences are affected by the inducement to use certain choice procedures.

The results of our experiment are threefold. First, inducing subjects to certain choice procedures has an effect on their welfare. Specifically, subjects' performance is distinctly better in the build and destroy treatments than in the baseline treatment, indicating that inducing subjects to use CBS heuristics (as opposed to holistic ones) increases their welfare. To the best of our knowledge, our paper is the first one to show that CBS procedures may be welfare increasing. If we interpret subjects' behaviour as being the implementation of some payoff-maximizing objective function subject to cognitive constraints, then our results suggest that inducing subjects to use CBS procedures relaxes their cognitive constraints and, as a result, improves their performance. We also detect a slight difference in performance between build and destroy in favour of the destroy treatment. We attribute such difference to the fact that, by the nature of the treatments themselves, it takes more time to construct an alternative in the build than in the destroy treatment. As a result, especially at relatively high time pressure levels, subjects tend to do better in the destroy treatment.

Second, by looking at disaggregated data by complexity, we find that the divergence in performance between baseline and build/destroy is maximised at intermediate complexity levels. That is, at simple problems the assignment to the treatment does not affect subjects' performance. As complexity increases the performance in the baseline treatment worsens at a higher rate than in the build/destroy treatment leading to the maximum degree of divergence at intermediate complexity levels. At relatively high complexity levels the performance across treatments tends to converge. This second finding suggests that the 'ecological rationality' of CBS heuristics holds *within a certain range of complexity*. At very simple and very complex problems subjects' performance is not affected by the nature of the heuristic induced. On the contrary, at moderately complex problems inducing CBS procedures, as opposed to holistic ones, pays off.

Third, we compare random choice (which we construct by running simulations) and with subjects' choice and find that subjects behaviour is distinctly different from random in both the build and destroy treatment. On the contrary, in the baseline treatment we find that subjects' choice is different from random choice at relatively simple

---

<sup>6</sup> Consider an experimental design similar to ours in which preferences are not controlled. Assume that we observe an experimental subject that—when faced with the same choice problem—makes different choices when induced to use different procedures. Two explanations are possible. First, the subject's preferences have not changed and one choice procedure has led him to make a better choice than some other choice procedure. Second, his preferences have changed and the experimental subject has chosen accordingly. Notice that not only by controlling preferences the second explanation is ruled out, but also the two explanations lead to very different conclusions: unlike in the second scenario, in the first one it makes sense to incentivize the use of the one of the two choice procedures (i.e., the welfare-improving one).

problems only. At relatively high complexity levels, we cannot rule random choice out. This result is consistent with the other findings.

The remainder of the paper is structured as follows: Section 2 presents the theoretical framework behind the experiment; Sect. 3 discusses the experimental design; Sect. 4 illustrates the results of the simulations; Sect. 5 presents the results of the experiment; Sect. 6 discusses the related literature and concludes. The supplementary material contains additional figures (including examples of screenshots), a detailed comparison of build and destroy treatments, and the instructions.

## 2 The theoretical framework

Our experiment can be described in terms of the choice-with-frames model proposed by Salant and Rubinstein (2008) that formalises and generalises the concept of framing effects (Tversky and Kahneman 1981). Let  $X$  denote a grand set of multi-attribute alternatives and  $\mathcal{S}$  a collection of non-empty subsets of  $X$ . Denote by  $\mathcal{R}$  a set of preference relations over  $X$ . Finally, denote by  $\mathcal{F}$  the set of ‘frames’—additional information other than the feasible set of alternatives that is irrelevant to a rational decision-maker, but might affect choices.<sup>7</sup> In the current setting we interpret a frame as an induced choice procedure, which can be either CBS or holistic. Hence, we let  $\mathcal{F} = \{\text{CBS}, \text{holistic}\}$ . For any  $\succ \in \mathcal{R}$ ,  $S \in \mathcal{S}$ , and  $f \in \mathcal{F}$ , we define  $c_{\succ}(S, f) : \mathcal{S} \rightarrow X$  with  $c_{\succ}(S, f) \in S$  as the experimental subject’s chosen alternative from the choice problem  $S$  under the induced preference  $\succ$  and the induced choice procedure  $f$ .

Our experimental design can be thought of as the triple  $\langle \mathcal{S}, \mathcal{R}, \mathcal{F} \rangle$  in the sense that we have full control over the set of choice problems, the preferences, which we induced via monetary incentives, and the induced choice procedure. Our main hypothesis is summarized in Eq. 1:

$$c_{\succ}(S, \text{CBS}) \stackrel{?}{\succ} c_{\succ}(S, \text{holistic}) \quad (1)$$

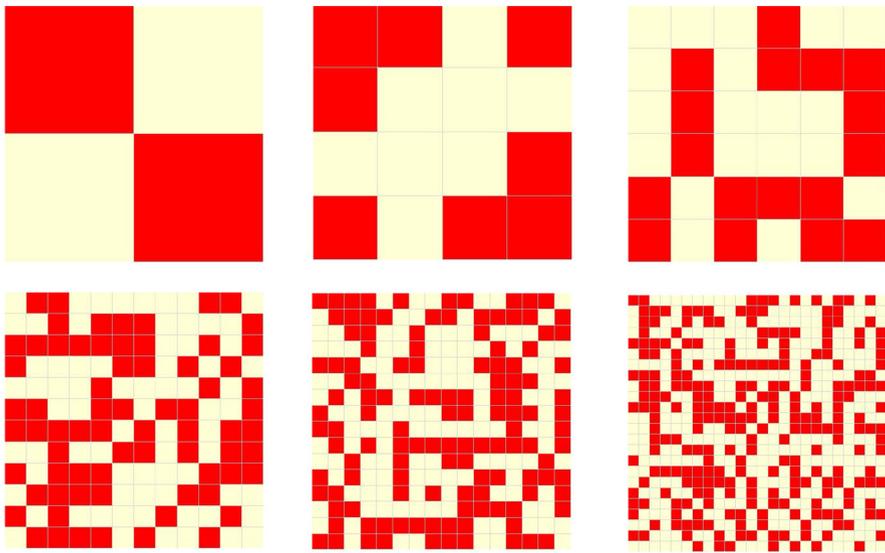
That is, we are interested in understanding whether—*ceteris paribus*—inducing subjects to use CBS vs. holistic procedures has an effect on their welfare, measured in terms of how high in the preference ranking the chosen alternative is.

## 3 Experimental design

### 3.1 The task

In all treatments of this experiment subjects are shown a target alternative—an abstract figure—and are financially incentivized to choose or construct the figure that most closely looks like the target. All figures in this experiment are grids of various dimensions, whose cells are coloured either red or beige (see Fig. 1). Our decision to use

<sup>7</sup> In the theoretical literature an example of frame is given by the ordering according to which alternatives are presented to the decisions-maker (Papi 2012).



**Fig. 1** Examples of figures of various complexity

abstract figures as alternatives is motivated by the fact that we wanted to abstract as much as possible from the context to collect generalizable results. Abstract figures of the kind considered here nicely serve this purpose, also because they can be partitioned into building blocks, which can naturally be interpreted as attributes or characteristics of the figures themselves. Throughout the paper we refer to each cell comprising a figure as *pixel* and to a set of one or more adjacent pixels as *block*.

In the baseline treatment, figures are given and subjects have to choose one among those available. In order to select a figure, subjects have to click on the figure they intend to select before the time expires and, as they do so, the selected figure appears enlarged next to the target.<sup>8</sup> The baseline treatment is meant to induce a holistic procedure, as alternatives are given and subjects are induced to make pairwise comparisons between the target and the selected alternative by inspecting all pixels of the selected alternative before exploring the next figure.

In the build and destroy treatments, on the other hand, figures are not given, but have to be constructed according to certain procedures. In the build treatment, subjects are shown several blocks and a figure partitioned into blank spots—which throughout the paper we refer to as *slots*—of the same dimensions as the blocks. Subjects are asked to construct a figure by placing the blocks in the slots within the time limit.<sup>9</sup> In order to insert a block into a slot, subjects have to first click on the block and then on the slot. Any block can be allocated to any slot (even to multiple slots) and can be replaced by any other block according to the procedure just described before the time expires.

<sup>8</sup> See figure 3 in the supplementary material for an example of a screenshot.

<sup>9</sup> See Fig. 4 in the supplementary material for an example of a screenshot.

In the destroy treatment, on the contrary, a figure partitioned into red blocks is presented to subjects, who have to change the colour (from red to beige) of the blocks.<sup>10</sup> In order to change the colour of a block, subjects have to click on the block they intend to change the colour to. The colour of a block can be repeatedly changed by replicating the above procedure before the time expires. For any given target, the blocks in the build and destroy treatment are different, but designed in such a way that the problems in the build and destroy treatments are exactly the same. We will come back to this in the next subsection. Both the build and the destroy treatments are meant to induce CBS procedures, as subjects are asked to construct what they think is their most preferred alternative by sequentially shaping its characteristics one by one.

The reason for which—besides the build treatment—we have introduced the destroy treatment is threefold. First, we wanted a second procedure other than the build treatment that induces a CBS heuristic. Second, constructive procedures, such as the build treatment, are not the only ones to induce CBS procedures. CBS procedures can be induced also via ‘destructive procedures’, such as the destroy treatment. Having both constructive and destructive procedures completes the analysis of CBS procedures and, as a result, increases the robustness of our results. Third, we wanted to check whether constructing a figure from scratch vs. decomposing an existing figure has an effect on subjects’ performance.

In all treatments once that subjects think they have selected (or constructed) the best alternative, they have to confirm their choice by clicking on the confirm button within the time limit. Subjects can replace the selected (or constructed) alternative as many times as they want before the time expires. If they select (or construct) an alternative without confirming within the time limit, the selected (or constructed) alternative (when the time expires) is considered to be their final choice. If at a choice problem subjects do not select (or construct) any alternative, their payoff is automatically set to be equal to zero as far as that choice problem is concerned.

In all treatments we count a mistake whenever a pixel of the selected (or constructed) figure is different from the corresponding pixel of the target figure. Let  $d_{\max}(c, t)$  (resp.,  $d_{\min}(c, t)$ ) denote the maximum (resp., minimum) number of mistakes a subject can commit at choice problem  $c$  in treatment  $t$ . Let  $d_i(c, t)$  denote the actual number of mistakes committed by subject  $i$  at choice problem  $c$  in treatment  $t$ . Subject  $i$ ’s payoff  $\pi_i(c, t)$  (in euros) at choice problem  $c$  in treatment  $t$  is defined as follows  $\pi_i(c, t) = 3 + 17 \cdot I_i(c, t)$ , where  $I_i(c, t) \equiv \frac{d_{\max}(c, t) - d_i(c, t)}{d_{\max}(c, t) - d_{\min}(c, t)}$  is the *performance index*. Hence, subjects earn a show-up fee of €3 plus a (linear) performance-based payment that lies in the interval [0, 17], which is inversely related with the number of mistakes a subject commits. Monetary payments are rounded to the closest euro cent.

### 3.2 Design of the choice problems

We varied the level of both time pressure, which can be interpreted as the search costs, and complexity. In particular, we set four different levels of time pressure—60, 80, 100, and 120 s—and six different levels of complexity, which we called Simple-

<sup>10</sup> See figure 5 in the supplementary material for an example of a screenshot.

**Table 1** Specification of the complexity levels

Compl.	# Pixel/alt.	# Alt.Ba.	# Alt.Bu.De.
SF	4	16	16
SC	16	16	16
MF	36	16	262,144
MC	144	16	262,144
DF	256	16	$18.4 \times 10^{18}$
DC	576	16	$18.4 \times 10^{18}$

Fine (SF), Simple-Coarse (SC), Medium-Fine (MF), Medium-Coarse (MC), Difficult-Fine (DF), and Difficult-Coarse (DC). In this experiment we measure complexity in terms of the fineness (i.e., number of pixels) of the figures that subjects have to work with (see Table 1). This definition of complexity is supported by the psychology literature, according to which visual complexity is related with a multiplicity of visual dimensions, such as ‘quantity of objects’, ‘symmetry’, and ‘variety of colours’ (Oliva et al. 2004). Figure 1 shows an example of a target alternative for each complexity level.

The second column of Table 1 reports the number of pixels that every figure of each complexity level is made of, which we arbitrarily decided.<sup>11</sup> The third column reports the number of alternatives available in the baseline treatment to choose from. The fourth column reports the number of distinct alternatives that can be constructed in the build and destroy treatment. Note that the ‘size’ of the choice problem grows exponentially with complexity in the build and destroy treatments. This has to do with the intrinsic nature of these two treatments, which we will explain below. Note that, on the contrary, the number of alternatives that we made available in the baseline treatment is equal to 16 regardless of the complexity level. Ideally, the size of each choice problem should be the same across all treatments. However, this was not possible, as it was unfeasible to present subjects with hundreds of thousands of figures to choose from in the baseline treatment. We addressed this issue by showing subjects 16 alternatives only in the baseline treatment that were extracted (without replacement) at random out of the set of alternatives that could potentially be constructed in the build and destroy treatments.

Table 2 provides details regarding the characteristics of the build (first three columns) and destroy treatments (fourth and fifth column): number of blocks per figure, number of slots per figure (for the build treatment only), and number of pixels per block. Recall that in the build treatment, subjects have to insert the blocks into the slots available. The number of alternatives that can be created is, therefore, given by the number of slots to the power of the number of slots. On the other hand, in the destroy treatments subjects are asked to change to colour of the blocks (either red or beige). Hence, the number of alternatives that can be generated in the destroy treatment is equal to two to the power of the number of blocks. By looking at Table 2,

<sup>11</sup> The only restriction was that the number of pixels has to have an integer square root to be able to construct a squared-shaped grid.

**Table 2** Comparison build vs. destroy

Compl.	Build			Destroy		# Alt.Bu.De.
	# Blocks	# Slots	# Pixel/block	# Blocks	# Pixel/block	
SF	2	4	1	4	1	16
SC	2	4	4	4	4	16
MF	4	9	4	18	2	262,144
MC	4	9	16	18	8	262,144
DF	16	16	16	64	4	$18.4 \times 10^{18}$
DC	16	16	36	64	9	$18.4 \times 10^{18}$

it can be seen that the size of the choice problems at each complexity level is the same across the build and destroy treatments. In section 2 of the supplementary material we graphically show that the sets of alternatives that can be created in these two treatments actually coincide.

In this experiment all target figures were constructed in the following way: Denote by  $x$  the number of pixels that form an alternative. We first created a squared-shaped figure made of  $x$  empty pixels. We then randomly extracted without replacement  $\frac{x}{2}$  pixels of that figure and coloured them red. We finally coloured the remaining pixels beige. We created one target alternatives for every combination of time pressure and complexity level. So overall we created 24 (4 time pressure times 6 complexity levels) plus an additional 6 (one for each complexity level for the practice rounds) target alternatives. For every combination of time pressure and complexity level, subjects faced exactly the same target alternative across the three treatments.

### 3.3 Implementation

A total of 58 experimental subjects were randomly recruited from a university database of undergraduates. Subjects were taken to the lab and shown the instructions.<sup>12</sup> Subsequently an experimenter read them loudly. Subjects were then asked to solve six practice rounds (one for each complexity level). At a later stage they were asked to solve six series (one for each complexity level) of 4 choice problems (one for each time-pressure level) that were valid for the calculation of their performance-based payment. The order of the series and the choice problems within each series were randomized as well as the order in which alternatives and blocks appeared on the screenshots.<sup>13</sup> After every round and at the end of each series feedback was given in terms of the actual number of mistakes relative to the minimum and maximum number

<sup>12</sup> See section 3 of the supplementary material for the instructions.

<sup>13</sup> Due to an unintended mistake in the software, the first series of choice problems in the baseline and build treatments was complexity Simple-Fine for all experimental subjects. This mistake did not seem to have affected the results.

of mistakes.<sup>14</sup> One choice problem out of 24 was selected at random for calculating the actual payoff. After the experiment subjects completed an anonymous questionnaire on demographics. We followed standard experimental procedures.

The design of the experiment is between-subject. Twenty subjects were assigned to the baseline treatment, twenty to the build, and eighteen to the destroy treatment. The experiment took place on the 4th of November 2014 (baseline and build) and on the 21st of April 2015 (destroy) at the Cognitive and Experimental Economics Laboratory (CEEL) of the University of Trento. The software used in the experiment was designed by the authors of the paper and the CEEL manager Mr Marco Tecilla.

## 4 Simulations

In order to implement the payoff function defined above, we needed to know the maximum  $d_{\max}(c, t)$  and the minimum  $d_{\min}(c, t)$  number of mistakes a subject can potentially commit at every choice problem  $c$  in each treatment  $t$ . While the calculation of these numbers is straightforward for the baseline treatment (the size of the choice problem is 16 regardless of complexity), figuring out  $d_{\max}(\cdot)$  and  $d_{\min}(\cdot)$  is not obvious as far as the build and destroy treatments are concerned, because hundreds of thousands of distinct alternatives can be constructed at relatively high complexity levels (see Table 1).

We addressed this issue by simulating random choice in the build treatment.<sup>15</sup> That is, we designed a software in which a ‘subject’ repeatedly draws at random (with replacement) a block from the sets of blocks available and inserts it into a slot, until a figure is completed. The resulting figure was considered to be the subject’s final choice. We then calculated the number of mistakes associated with it. We iterated this procedure 500,000 times for every target alternative that we used in this experiment. The results of the simulations are shown in Table 3, reporting the estimates of  $d_{\max}(\cdot)$ ,  $d_{\min}(\cdot)$ , and the average number of mistakes.<sup>16</sup>

Note that choices are non-trivial for most complexity levels. The perfect fit between target and constructed alternative can be reached only at the SF complexity level, which is meant to be a rationality check. As complexity increases, the minimum and maximum number of mistakes a subject can commit increases. At the highest

<sup>14</sup> See figure 6 in the supplementary material for an example.

<sup>15</sup> Since the build and destroy treatments are equivalent, the results of the simulations hold also for the destroy treatment.

<sup>16</sup> Since the values of  $d_{\max}(\cdot)$  and  $d_{\min}(\cdot)$  recovered from the simulations are only estimates of their respective true values, then we slightly modified the payoff function defined above as follows:

$$\pi_i(c, t) \equiv \begin{cases} 3 & \text{if } d_i(c, t) > d_{\max}(c, t) \\ 3 + 17 \cdot I_i(c, t) & \text{if } d_i(c, t) \in [d_{\min}(c, t), d_{\max}(c, t)] \\ 20 & \text{otherwise} \end{cases}$$

**Table 3** Results of the simulations: estimates of  $d_{\max}^c$ ,  $d_{\min}^c$ , and average number of mistakes

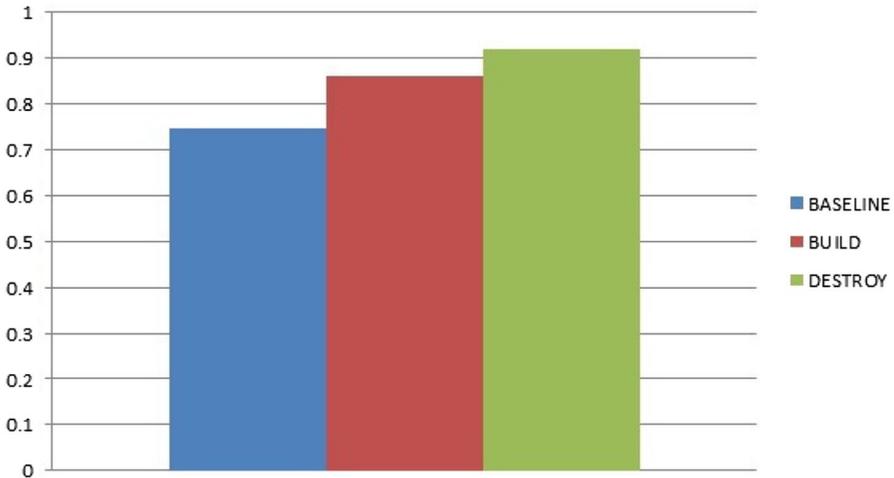
Compl.	$d_{\max}(\cdot)$	$d_{\min}(\cdot)$	Average
SF	4	0	2
SC	12	4	8
MF-60s	25	11	18
MF-80s	27	9	18
MF-100s	26	10	18
MF-120s	28	8	18
MC-60s	91	53	72
MC-80s	88	56	72
MC-100s	90	54	72
MC-120s	85	59	72
DF-60s	163	93	128
DF-80s	166	84	128
DF-100s	164	92	128
DF-120s	163	87	128
DC-60s	339	236	288
DC-80s	339	236	288
DC-100s	335	241	288
DC-120s	334	243	288

complexity level, the best fit involves about 240 differences in pixels between target and constructed alternative.<sup>17</sup>

By plotting the distribution of mistakes of subjects choosing random, we found that it looks very similar to a normal distribution.<sup>18</sup> That is, the distribution is symmetric around the average number of mistakes and the mean, the median, and the mode coincide. This observation is important, because it reveals that a subject choosing at random would get most of the times a performance index equal to 0.5. This result follows from the fact that the distribution of the figures constructed by using the above method is multinomial (recall that a block is drawn at random with replacement from a set of two or more blocks and inserted into an empty slot until a figure is completed). Since the number of mistakes is calculated by comparing every pixel of the constructed figures with the corresponding pixel of the target figure, then, as the sample size increases, the distribution of the number of mistakes converges towards a normal distribution. We will use again the results of the simulations in a section below to compare subjects' behaviour with random choice.

<sup>17</sup> We reported only one entry of the maximum, minimum, and average number of mistakes for both the SC and the SF complexity levels. This because at those complexity levels the composition of target alternative does not change the distribution of mistakes.

<sup>18</sup> See figures 7 and 8 in the supplementary material.



**Fig. 2** Average performance—pooled data

## 5 Results

In this section we measure subjects' performance by using the performance index  $I_i(c, t)$  that determines the performance-based payment, which we previously introduced. Recall that this index is linear, lies in the interval  $[0, 1]$ , and is inversely related with the number of mistakes a subject commits.

### 5.1 Between-subject analysis

Pooled results are shown in Fig. 2. The treatment in which subjects performed the best is the destroy, followed by the build and the baseline. However, while the difference in performance between baseline and build (and destroy) is relatively substantial, the difference between build and destroy is smaller.

We checked whether the difference in performance across treatments is statistically significant by performing a Mann–Whitney and a Kolmogorov–Smirnov test.<sup>19</sup> For every subject  $i$ , we averaged the performance index across complexity levels and across time-pressure levels and then compared the averages across treatments. The results—summarised in Table 4—suggest that the difference in performance between all treatments is significant at the 1% level.

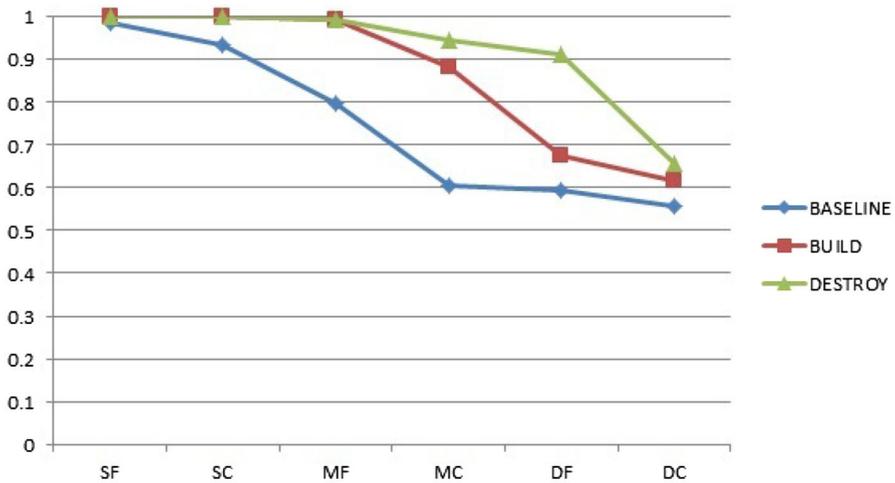
By disaggregating by complexity, one can see that the difference in performance is maximised for intermediate levels of complexity (see Fig. 3). In particular, when the choice problem is very simple, the allocation to the treatment does not affect subject's performance. As complexity increases, subject's performance progressively deteriorates in all treatments. However, the extent to which it worsens differs across

<sup>19</sup> The reason for which we used both the Mann–Whitney and the Kolmogorov–Smirnov test is because the equality of variances assumption—needed for the Mann–Whitney test to work—does not always hold.

**Table 4** Difference in average performance (pooled data)—baseline ( $n = 20$ ) vs. build ( $n = 20$ ) vs. destroy ( $n = 18$ )

Compl.	Mann–Whitney		Kolmogorov–Smirnov	
	St. Test Stat.	Exact. Sign.	Test Stat.	Asympt. Sig.
Baseline vs. build	4.923	0.000***	2.688	0.000***
Baseline vs. destroy	5.262	0.000***	3.078	0.000***
Build vs. destroy	3.245	0.001***	1.984	0.001***

\*\*\* 1% significance

**Fig. 3** Average performance by complexity

treatments. Performance in the baseline gets worse relatively soon, followed by that of the build treatment. The performance in the destroy treatment is the last one to deteriorate. As complexity increases to the highest levels, subject's performance across treatments tends to converge.

We performed a Mann–Whitney and a Kolmogorov–Smirnov test to check whether the difference in the distribution of the performance index across treatments is statistically significant (see Tables 5, 6, and 7). For every subject  $i$  and treatment  $t$ , we averaged the performance index across time-pressure levels for each complexity level and then compared the averages across treatments. By comparing the baseline and the build treatment, we found that the difference is significant at 1% level for the complexity levels SC, MF, and MC and insignificant at the other complexity levels. We detected a similar pattern when we compared baseline and destroy. On the contrary, the difference between build and destroy is significant (at the 1%) only for the complexity level DF.

Figure 4 shows subjects' performance disaggregated by time pressure. Subjects did better in the destroy than in the build treatment for every level of time pressure. Average performance in the build treatment is similar to that of the baseline for high

**Table 5** Difference in average performance by complexity—baseline ( $n = 20$ ) vs. build ( $n = 20$ )

Compl.	Mann–Whitney		Kolmogorov–Smirnov	
	St.Test Stat.	Exact.Sign.	Test Stat.	Asympt.Sig.
SF	2.082	0.289	0.632	0.819
SC	4.022	0.001***	1.897	0.001***
MF	5.510	0.000***	3.004	0.000***
MC	5.064	0.000***	2.846	0.000***
DF	1.785	0.076*	1.107	0.172
DC	1.163	0.253	0.791	0.560

\* 10% significance; \*\*\* 1% significance

**Table 6** Difference in average performance by complexity—baseline ( $n = 20$ ) vs. destroy ( $n = 18$ )

Compl.	Mann–Whitney		Kolmogorov–Smirnov	
	St.Test Stat.	Exact.Sign.	Test Stat.	Asympt.Sig.
SF	1.979	0.303	0.616	0.843
SC	3.848	0.001***	1.847	0.002***
MF	5.312	0.000***	2.907	0.000***
MC	5.274	0.000***	3.078	0.000***
DF	5.263	0.000***	3.078	0.000***
DC	2.222	0.026**	1.129	0.156

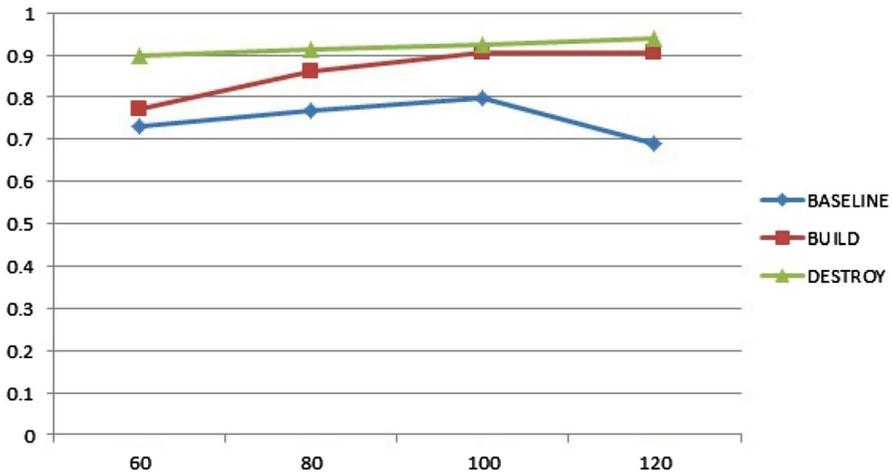
\*\* 5% significance; \*\*\* 1% significance

**Table 7** Difference in average performance by complexity—build ( $n = 20$ ) vs. destroy ( $n = 18$ )

Compl.	Mann–Whitney		Kolmogorov–Smirnov	
	St.Test Stat.	Exact.Sign.	Test Stat.	Asympt.Sig.
SF	0.000	1.000	0.000	1.000
SC	0.000	1.000	0.000	1.000
MF	−0.082	0.965	0.154	1.000
MC	2.153	0.030**	1.146	0.145
DF	4.444	0.000***	2.445	0.000***
DC	0.906	0.377	0.735	0.652

\*\* 5% significance; \*\*\* 1% significance

time-pressure levels. As time pressure is relaxed, average performance in the baseline treatment increases and converges towards that of the destroy treatment. Interestingly, performance in the baseline treatment decreases by relaxing time pressure from 100 to 120 s. There are two possible explanations. First, subjects got bored and started to play randomly. We will explicitly analyse random choice in the next subsection. Second, whenever subjects have relatively little time (or, equivalently, the search cost is high), then they look for a local optimum and, as soon as they discover one, they



**Fig. 4** Average performance by time pressure

**Table 8** Difference in average performance by time pressure—baseline ( $n = 20$ ) vs. build ( $n = 20$ )

Time Pr. (s)	Mann–Whitney		Kolmogorov–Smirnov	
	St.Test Stat.	Exact.Sign.	Test Stat.	Asympt.Sig.
60	0.893	0.383	0.791	0.560
80	3.435	0.000***	1.423	0.035**
100	4.355	0.000***	2.530	0.000***
120	5.302	0.000***	2.846	0.000***

\*\* 5% significance; \*\*\* 1% significance

stop searching. On the other hand, when subjects realize that they have relatively more time, they keep searching, despite having possibly identified a local optimum, because their goal is to identify the global optimum. However, by doing so they end up selecting a worse alternative.

In order to perform Mann–Whitney and Kolmogorov–Smirnov, we averaged—for every subject and treatment—the performance index across complexity levels for each time-pressure level and then compared the averages across treatments. The tests suggest that the difference in the distribution of the performance index between baseline and build is statistically significant at 1% level for 80, 100, and 120 s time-pressure levels. On the other hand, the divergence in performance between baseline and destroy is significant for all time-pressure levels. By comparing the build and the destroy treatments, the difference is statistically significant only for a time pressure of 60 s. At the other time-pressure levels, the difference is either insignificant or significant at higher levels. See Tables 8, 9, and 10 for the details.<sup>20</sup>

<sup>20</sup> In the destroy treatment we implicitly set a lower bound on the performance index, because by changing the colour of one block only, subjects can complete a figure. By doing so, subjects would get a performance index equal to approximately 0.5, because, as explained above, we constructed the target alternatives by randomly colouring half of the pixels red and the remaining half beige. A subject playing in this way would

**Table 9** Difference in average performance by time pressure—baseline ( $n = 20$ ) vs. destroy ( $n = 18$ )

Time Pr. (s)	Mann–Whitney		Kolmogorov–Smirnov	
	St.Test Stat.	Exact.Sign.	Test Stat.	Asympt.Sig.
60	4.970	0.000***	2.616	0.000***
80	4.736	0.000***	2.411	0.000***
100	4.941	0.000***	2.599	0.000***
120	5.264	0.000***	3.078	0.000***

\*\*\* 1% significance

**Table 10** Difference in average performance by time pressure—build ( $n = 20$ ) vs. destroy ( $n = 18$ )

Time Pr. (s)	Mann–Whitney		Kolmogorov–Smirnov	
	St.Test Stat.	Exact.Sign.	Test Stat.	Asympt.Sig.
60	3.128	0.001***	2.001	0.001***
80	2.310	0.020**	1.300	0.068*
100	1.257	0.217	1.146	0.145*
120	1.785	0.077*	1.060	0.211

\* 10% significance; \*\* 5% significance; \*\*\* 1% significance

We also looked at whether demographics have an effect on subjects' performance. We could not find any systematic difference in behaviour across demographic groups both within and across treatments.

## 5.2 Random choice

In order to investigate whether subjects' choice is different from random choice, we first generated random choice data. In the baseline treatment this task was straightforward, as subjects chose among 16 alternatives only. Hence, we assigned a  $\frac{1}{16}$  probability to every option at each choice problem. In the build and destroy treatments, we used the results of the simulations that we discussed in Sect. 4.

Once that random choice was generated, we identified the maximum and minimum number of mistakes a subject could commit and the median number of mistakes for

---

Footnote 20 continued

behave in a way that is analogous to a subject choosing randomly (see Sect. 4). However, the data clearly reveal that such lower bound is not binding for at least two reasons. First, the average number of moves (defined as the number of times a subject changed the colour of a block) in the destroy treatment is greater than one. Specifically, the average number of moves (with standard deviation in brackets) is 2.15 (0.32) at the SF, 2.13 (0.19) at the SC, 6.71 (2.70) at the MF, 7.71 (1.96) at the MC, 19.96 (6.85) at the DF, and 24.26 (9.99) at the DC complexity levels. In terms of time pressure, the average number of moves (with standard deviation in brackets) is 9.62 (3.01) at 60 s, 10.65 (3.10) at 80s, 10.16 (3.44) at 100 s, and 11.52 (3.75) at 120 s. Second, as it will become clearer in the next subsection, subjects clearly chose differently from random in the destroy treatment.

**Table 11** Subject vs. random choice: baseline treatment ( $n = 20$ )

Target		Binomial			Wilcoxon Signed Rank		
Compl.	Time Pr. (s)	St.Test Stat.	$p$ value	Sig.	St.Test Stat.	$p$ value	Sig.
SF	60	-4.249	0.000	***	4.472	0.000	***
SF	80	-4.249	0.000	***	4.472	0.000	***
SF	100	-4.249	0.000	***	4.300	0.000	***
SF	120	-4.249	0.000	***	4.300	0.000	***
SC	60	-4.249	0.000	***	4.379	0.000	***
SC	80	-3.801	0.000	***	3.962	0.006	***
SC	100	-4.249	0.000	***	4.179	0.000	***
SC	120	-4.249	0.000	**	4.053	0.000	***
MF	60	-2.460	0.012	**	2.298	0.022	**
MF	80	-3.801	0.000	***	3.948	0.000	***
MF	100	-3.354	0.000	***	4.009	0.000	***
MF	120	0.000	1.000	-	2.712	0.007	***
MC	60	-1.565	0.115	-	1.990	0.047	**
MC	80	-2.460	0.012	**	3.302	0.001	***
MC	100	-2.012	0.041	**	2.419	0.016	**
MC	120	-1.118	0.263	-	0.678	0.498	-
DF	60	-0.671	0.503	-	2.103	0.035	**
DF	80	0.671	0.503	-	0.113	0.910	-
DF	100	-2.460	0.012	**	2.908	0.004	**
DF	120	-0.671	0.503	-	0.300	0.764	-
DC	60	0.000	1.000	-	-0.975	0.329	-
DC	80	-0.224	0.824	-	1.653	0.098	*
DC	100	-0.224	0.824	-	0.824	0.410	-
DC	120	-2.182	0.027	**	2.938	0.003	***

\* 10% significance; \*\* 5% significance; \*\*\* 1% significance

each choice problem. We then calculated the performance index associated with the median choice. Finally, we tested whether the difference in the performance index associated with the median random choice is statistically different from the performance index associated with the subjects' median choice by using the binomial and the Wilcoxon signed rank tests. Detailed results are reported in Tables 11, 12, and 13.

In the baseline treatment subjects distinctly chose different from random at relatively simple problems. We indeed find that the difference between subjects' and random median choice is statistically significant at 5% (or below). At relatively harder problems, we do not detect any statistically significant difference. On the contrary, in the build and destroy treatments the difference between subjects' and random median choice is statistically significant at almost all choice problems regardless of complexity.

Interestingly, we record that in the build treatment subjects' choice is not statistically different from random at difficult problems (DF and DC) with very high time pressure (60 s). Recall that at the DF and DC complexity levels subjects have to fill in 16 slots to

**Table 12** Subject vs. random choice: build treatment ( $n = 20$ )

Target		Binomial			Wilcoxon Signed Rank		
Compl.	Time Pr. (s)	St.Test Stat.	$p$ value	Sig.	St.Test Stat.	$p$ value	Sig.
SF	60	-3.928	0.000	***	4.472	0.000	***
SF	80	-3.928	0.000	***	4.472	0.000	***
SF	100	-3.928	0.000	***	4.472	0.000	***
SF	120	-3.928	0.000	***	4.472	0.000	***
SC	60	-3.928	0.000	***	4.472	0.000	***
SC	80	-3.928	0.000	***	4.472	0.000	***
SC	100	-3.928	0.000	***	4.472	0.000	***
SC	120	-3.928	0.000	***	4.472	0.000	***
MF	60	-3.928	0.000	***	4.308	0.000	***
MF	80	-4.364	0.000	***	4.412	0.000	***
MF	100	-3.928	0.000	***	4.379	0.000	***
MF	120	-3.928	0.000	***	4.379	0.000	***
MC	60	-3.055	0.001	***	2.561	0.010	***
MC	80	-3.928	0.000	***	3.302	0.001	***
MC	100	-3.928	0.000	***	2.419	0.016	**
MC	120	-4.364	0.000	***	4.040	0.000	***
DF	60	0.436	0.664	-	-1.596	0.110	-
DF	80	-2.619	0.007	***	2.557	0.011	**
DF	100	-3.055	0.001	***	3.189	0.001	***
DF	120	-3.491	0.000	***	3.272	0.001	***
DC	60	-0.873	0.383	-	-0.122	0.903	-
DC	80	-2.182	0.027	**	1.791	0.073	*
DC	100	-3.055	0.001	***	3.684	0.000	***
DC	120	-2.294	0.019	**	3.141	0.002	***

\* 10% significance; \*\* 5% significance; \*\*\* 1% significance

make a figure, which is a time-consuming task. By looking at the data, we found that at very hard problems under high time pressure many subjects did not complete a figure within the time limit. Hence, the reason for which we observe such pattern is that, as specified in the instructions, a performance-based payoff equal to zero is assigned if subjects fail to select (or construct) a figure within the time limit. As expected, we do not detect such pattern in the build treatment.

## 6 Related literature and conclusion

Our paper is interdisciplinary and related with the multi-attribute individual decision-making literature in both psychology and economics. Daniel Kahneman and Amos Tversky pioneered a fruitful tradition in psychology on multi-attribute individual decision-making by establishing the so-called ‘heuristic-and-biases’ research agenda

**Table 13** Subject vs. random choice: destroy treatment ( $n = 18$ )

Target		Binomial			Wilcoxon Signed Rank		
Compl.	Time Pr. (s)	St.Test Stat.	$p$ value	Sig.	St.Test Stat.	$p$ value	Sig.
SF	60	-3.671	0.000	***	4.243	0.000	***
SF	80	-3.671	0.000	***	4.243	0.000	***
SF	100	-3.671	0.000	***	4.243	0.000	***
SF	120	-3.671	0.000	***	4.243	0.000	***
SC	60	-3.671	0.000	***	4.243	0.000	***
SC	80	-3.671	0.000	***	4.243	0.000	***
SC	100	-3.671	0.000	***	4.243	0.000	***
SC	120	-3.671	0.000	**	4.243	0.000	***
MF	60	-3.671	0.000	***	4.140	0.000	***
MF	80	-4.129	0.000	***	4.062	0.000	***
MF	100	-3.671	0.000	***	4.243	0.000	***
MF	120	-3.671	0.000	***	4.146	0.000	***
MC	60	-3.671	0.000	***	3.838	0.000	***
MC	80	-3.671	0.000	***	3.789	0.000	***
MC	100	-3.671	0.000	***	3.859	0.000	***
MC	120	-4.129	0.000	***	3.894	0.000	***
DF	60	-3.671	0.000	***	3.389	0.000	***
DF	80	-3.671	0.000	***	3.790	0.000	***
DF	100	-3.671	0.000	***	3.780	0.000	***
DF	120	-3.671	0.000	***	3.870	0.000	***
DC	60	-2.294	0.019	**	2.657	0.008	***
DC	80	-2.753	0.004	**	3.340	0.001	***
DC	100	-1.835	0.064	*	2.680	0.007	***
DC	120	-2.012	0.041	**	0.937	0.349	-

\* 10% significance; \*\* 5% significance; \*\*\* 1% significance

(Gilovich et al. 2002). Given that humans are irrational, its goal is to investigate the heuristics that individuals use to make decisions and discover potential biases in their behaviour. As an example, a framing effect refers to the phenomenon whereby individuals change their decision whenever the same choice problem is presented in different ways (Tversky and Kahneman 1981). Unlike the heuristic-and-biases tradition that mainly focuses on decisions under uncertainty, Payne et al. (1993) uses verbal protocols and mouselab to test adaptivity in multi-attribute decision-making under certainty. They find that people adapt the choice procedure to the choice environment and are willing to save cognitive effort. In particular, subjects tend to use holistic procedures when more weight is put on the goal of maximizing accuracy with respect to the goal of minimizing cognitive effort. On the other hand, under time pressure decision-makers seem to use CBS heuristics. The ‘fast-and-frugal’ research tradition further develops the work of Payne et al. (1993) by—unlike the heuristic-and-biases approach -investigating the conditions under which a certain class of heuristics better

describes subjects' behaviour (Gigerenzer et al. 1999; Gigerenzer and Selten 2001).<sup>21</sup> For example, Rieskamp and Hoffrage (2008) use the mouselab technique to investigate how inference strategies are affected by the amount of time pressure. They find that under relatively high (resp., low) time pressure, subjects tend to use CBS (resp., holistic) heuristics, such as the lexicographic procedure. More recently, the fast-and-frugal approach has increased its attention to real-world applications (Gigerenzer et al. 2011; Gigerenzer 2015).

The appearance of CBS procedures in economics goes back at least to Rubinstein (1988) that implicitly proposes a CBS model to explain the Allais paradox by assuming that the decision-maker assesses prizes and probabilities separately when choosing among lotteries. A first important experimental-economics contribution to this literature is Gabaix et al. (2006) that uses mouselab test individual's information acquisition patterns in an  $N$ -good game. That is, subjects are presented with an  $N \cdot M$  matrix of numbers. Each row is interpreted as an alternative and each column as an attribute. Subjects have to choose the row that maximises the algebraic sum of the numbers arranged on the row. Gabaix et al. (2006) find that the directed cognition model—according to which subjects perform the next search operation as if it was the last one—predicts the aggregate information acquisition patterns that subjects follow.<sup>22</sup> Unlike Gabaix et al. (2006), Reutskaja et al. (2011) use the more sophisticated eye-tracking technique to investigate consumers' search and choice attitudes on snack items under very high time pressure. They find that subjects' behaviour is consistent with a hybrid of the standard and the satisficing model of search.<sup>23</sup> The closest study to our work is Arieli et al. (2011) that also uses eye-tracking to test whether subjects use holistic or CBS procedures while choosing binary lotteries. Arieli et al. (2011) find that whenever the computation of the expected value is difficult, subject's eye movements are consistent with the use of CBS heuristics. On the contrary, when computations are easier, subjects' behaviour is consistent with a hybrid of a CBS and holistic procedure.

Unlike our work, the above branches of literature investigate whether a certain heuristic (or class of heuristics) describes subjects' behaviour. On the contrary, motivated by the proliferation of online choice platforms, we examine whether inducing subjects to use certain heuristics has an effect on their choice behaviour by proposing an innovative experiment. We are the first to show that CBS procedures may be welfare increasing. In particular, our results suggest that at intermediate levels of complexity inducing subjects to use CBS procedures (as opposed to holistic ones) makes them better off. On the contrary, at very simple and very complex problems the induced

<sup>21</sup> The main difference between Payne et al. (1993) and the fast-and-frugal approach is the definition of accuracy of an heuristic. While Payne et al. (1993) define it in terms of how different is the heuristic employed by a decision-maker from the optimal one, the fast-and frugal approach interprets it as how well the heuristic used by a decision-maker describes subjects' behaviour.

<sup>22</sup> More recently, Sanjurjo (2015) uses Gabaix et al. (2006)'s data to test his multi-attribute optimal search model. He finds that subjects investigate too much within alternatives and fail to optimally select the next alternatives to be explored.

<sup>23</sup> See also Caplin et al. (2011) for an experiment testing the satisficing model of search. Another study that uses eye-tracking is Shi et al. (2013), according to which information acquisition patterns in a multi-attribute online choice environment are analysed. The authors find that subjects frequently change search strategies and look up a few attributes or alternatives before switching strategy.

heuristic does not affect the subjects' performance. Finally, we show that when subjects are encouraged to use a holistic procedure, we cannot rule out the hypothesis that their choice behaviour is equivalent to that generated by random choice at relatively complex problems.

Our results provide support the growing theoretical economic literature that assumes individuals to use CBS choice procedures, such as shortlisting (Manzini and Mariotti 2007; Apestequia and Ballester 2013).<sup>24</sup> More broadly, our results support Herbert Simon (Simon 1955, 1956)'s intuition that the choice environment plays a crucial role in determining which heuristic decision-makers use to make decisions. We hypothesise that the reason for which CBS heuristics are more efficient than holistic ones is two-fold. On the one hand, CBS procedures encompass a more natural cognitive process than holistic ones. On the other hand, unlike holistic procedures, CBS procedures decompose difficult problems into simpler ones and, as a result, allow decision-makers to achieve better results at relatively harder problems. In order to test this conjecture we intend to run a follow-up experiment in which—after experimenting—subjects are asked to choose both the choice platform (either holistic or CBS) and the complexity level, where incentives are such that to harder problems correspond a higher monetary reward, *ceteris paribus*. We expect subjects preferring 'CBS choice platforms' to choose relatively high complexity levels.<sup>25</sup>

We view this paper as a first step to investigate the broad research question that we propose. Our experimental design is flexible and can naturally be extended in multiple intriguing directions. First, we provide evidence that as time pressure is relaxed, performance improves. Moreover, we show that at very complex problems whether subjects are induced to use holistic or CBS procedures does not have a significant effect on their performance. A first robustness check would be to verify what happens if (i) there is no time limit and (ii) incentives are increased at complex problems.

Second, we measured complexity of a figure with the number of pixels it is made of, which is consistent with the psychology literature (Oliva et al. 2004). We are aware that there are alternative measures of complexity that we could have used. For example, we have looked at the computer science literature and found a metric for complexity—called Kolmogorov complexity—based on algorithmic information theory. It turns out that a visual form can be represented by a string of symbols (Donderi 2006). In our experiment figures can be seen as matrices of zeros and ones and can, therefore, be described as strings. Informally, the Kolmogorov complexity of a string is defined as the length of the shortest program needed to reproduce the string itself

---

<sup>24</sup> See also Bernheim and Rangel (2007), Rubinstein and Salant (2012), and Apestequia and Ballester (2015) as decision theoretic papers that propose welfare measure under the assumption that decision-makers are not necessarily fully rational.

<sup>25</sup> Our results also provide support to the recognition-by-components (RBC) theory (Biederman 1987). The RBC theory is a bottom-up theory of visual perception according to which humans recognise an object by recognising its 'geons'—i.e., primitive geometric figures of which the object is made, such as blocks, cylinders, edges, etc. (Sternberg and Sternberg 2012). While the build and destroy treatment explicitly partition a figure into primitive components, the baseline treatment does not do so and, as a result, complicates the recognition by components. The fact that inducing subjects to use CBS as opposed to holistic procedures makes them better off indicates that a process of recognition by components facilitates the subjects' task in this experiment.

(Li and Vitányi 1997). For example, the 20-digit string 101010101010101010 is Kolmogorov-simpler than the 20-digit string 11010000111010111111, as while the shortest way of describing the former is to write it as a 10-time repetition of the binary string 10, the shortest way of describing the latter is to fully rewrite it, as there are no patterns in it. An interesting extension to the current experimental design would be to generate target figures of different Kolmogorov complexity and verify whether our results are robust to the modification.

Third, in this experiment we used abstract figures as objects of choice because we wanted a frame that was as neutral as possible in order to collect generalizable results. In a follow-up experiment, instead of abstract figures, we could use lotteries (Arieli et al. 2011), algebraic sums (Payne et al. 1993; Gabaix et al. 2006; Caplin et al. 2011) or real goods (Reutskaja et al. 2011) as objects of choice and compare the results with other studies.

Fourth, a feature of the current experiment is that no tradeoffs between attributes are generated in the build (and destroy) treatment. That is, it is not the case that by constructing a figure with certain characteristics (i.e., allocating a certain bloc in a certain slot) subjects are restricted in shaping the other characteristics. Instead, subjects have to identify ‘their most preferred attribute value’ for each attribute independently of any other attribute. In the real world, on the contrary, there are examples of both conflicting and non-conflicting attributes. An example of the former is given by the fact that if an individual—searching for a flat through a web-search engine—increases (resp., decreases) the number of desired bedrooms, then the rent of the shortlisted flats necessarily increases (resp., decreases), other things being equal. As an example of the latter, the fact that one chooses a red colour for the Ferrari through the car configurator does not restrict at all the choice of the wheel cap. One way of investigating whether our results are robust to the heterogeneity in the presence/absence of tradeoffs between attributes is to use real goods as alternatives.

Fifth, in this experiment we examined the effects of encouraging subjects to use certain classes of heuristics by inducing preferences, because we wanted to start from a scenario over which we have full control. This has given us a first insight into how *choices* are affected by inducing individuals to use certain choice procedures, *given preferences*. The natural following step is to investigate whether inducing subjects to use certain heuristics has an effect on *preferences* as well. We plan to investigate these issues in a series of follow-up experiments.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Apestequia, J., & Ballester, M. A. (2013). Choice by sequential procedures. *Games and Economic Behavior*, 77, 90–99.
- Apestequia, J., & Ballester, M. A. (2015). A measure of rationality and welfare. *Journal of Political Economy*, 77, 90–99.

- Arieli, A., Ben-Ami, Y., & Rubinstein, A. (2011). Tracking decision-makers under uncertainty. *American Economic Journal: Microeconomics*, 3(4), 68–76.
- Bernheim, D., & Rangel, A. (2007). Toward choice-theoretic foundations for behavioral welfare economics. *American Economic Review Paper and Proceedings*, 97(2), 464–470.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2), 115–147.
- Camerer, C., & Hogarth, R. (1999). The effects of financial incentives in economics experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 18, 7–42.
- Caplin, A., & Dean, M. (2011). Search, choice, and revealed preference. *Theoretical Economics*, 6, 19–48.
- Caplin, A., Dean, M., & Martin, D. (2011). Search and satisficing. *American Economic Review*, 101(7), 2899–2922.
- Donderi, D. C. (2006). Visual complexity: A review. *Psychological Bulletin*, 132(1), 73–97.
- Gabaix, X., Laibson, D., Moloche, G., & Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, 96(4), 1043–1068.
- Gigerenzer, G. (2015). *Simply Rational. Decision-Making in the Real World*. Oxford: Oxford University Press.
- Gigerenzer, G., & Selten, R. (2001). *Bounded Rationality: The Adaptive Toolbox*. Cambridge: MIT Press.
- Gigerenzer, G., Todd, P. M. & the ABC Research Group (1999). *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.
- Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics. The Foundations of Adaptive Behavior*. Oxford: Oxford University Press.
- Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and Biases. The Psychology of Intuitive Judgment*. Cambridge: Cambridge University Press.
- Li, M., & Vitányi, P. (1997). *An Introduction to Kolmogorov Complexity and Its Applications* (2nd ed.). Berlin: Springer.
- Manzini, P., & Mariotti, M. (2007). Sequentially rationalizable choice. *American Economic Review*, 97(5), 1824–1839.
- Oliva, A., Mack, M. L., Shrestha, M., Peeper, A. (2004). Identifying the perceptual dimensions of visual complexity of scenes. In *Proceeding of the 26th Annual Cognitive Science Society* (pp. 1041–1046)
- Papi, M. (2012). Satisficing choice procedures. *Journal of Economic Behavior and Organization*, 84, 451–462.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision Maker*. Cambridge: Cambridge University Press.
- Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *American Economic Review*, 101, 900–926.
- Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica*, 127, 258–279.
- Rubinstein, A. (1988). Similarity and decision-making under risk (is there a utility theory resolution to the allais paradox?). *Journal of Economic Theory*, 46(1), 145–153.
- Rubinstein, A., & Salant, Y. (2006). A model of choice from lists. *Theoretical Economics*, 1(1), 3–17.
- Rubinstein, A., & Salant, Y. (2012). Eliciting welfare preferences from behavioral datasets. *Review of Economic Studies*, 79, 375–387.
- Salant, Y., & Rubinstein, A. (2008).  $(a, f)$ : Choices with frames. *Review of Economic Studies*, 75(4), 1287–1296.
- Sanjurjo, A. (2015). Search with multiple attributes: Theory and empirics. *mimeo*.
- Shi, S. W., Wedel, M., & Pieters, F. G. M. R. (2013). Information acquisition during online decision-making: A model-based exploration using eye-tracking data. *Management Science*, 59(5), 1009–1026.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1), 99–118.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2), 129–138.
- Smith, V. L. (1976). Experimental economics: Induced value theory. *American Economic Review*, 66(2), 274–279.
- Sternberg, R. J., Sternberg, K. (2012). *Cognitive Psychology* (6th ed.). CENGAGE Learning.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79(4), 281–299.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453–458.