# Testing Models of Complexity Aversion

Konstantinos Georgalos* Nathan Nabil†

January 16, 2025

**Abstract**

In this study we aim to test behavioural models of complexity aversion. In this framework, complexity is defined as a function of the number of outcomes in a lottery. Using Bayesian inference techniques, we re-analyse data from a lottery-choice experiment. We quantitatively specify and estimate adaptive toolbox models of cognition, which we rigorously test against popular expectation-based models; modified to account for complexity aversion. We find that for the majority of the subjects, a toolbox model performs best both in-sample, and with regards to its predictive capacity out-of-sample, suggesting that individuals resort to heuristics in the presense of extreme complexity.

*Keywords*: Complexity aversion · Toolbox models · Heuristics · Risky choice · Bayesian modelling

*JEL codes:* C91 · D81 · D91

*Department of Economics, Lancaster University Management School, LA1 4YX, Lancaster, U.K.,✉: k.georgalos@lancaster.ac.uk, ORCID: 0000-0002-2655-4226

†Department of Economics, Lancaster University Management School, LA1 4YX, Lancaster, U.K.
✉ n.nabil@lancaster.ac.uk
We are grateful to Peter Moffatt, Stefania Sitzia and Daniel Zizzo for making their data available, and to Daniel Zizzo for providing useful comments. The support of The High End Computing (HEC) facility at Lancaster University is gratefully recognised.

# 1 Intro

In the recent years, the economic environment has witnessed a noticeable surge in complexity, driven by a confluence of interconnected factors. Technological advancements and globalization have expanded choices and convenience, while at the same time they have introduced overwhelming options that demand more of the consumers' attention and time. Mortgages, financial products, investment decisions and cryptocurrencies, all come with a plethora of options and features, that can exacerbate consumer decision-making, contributing to their increased cognitive fatigue.

In the field of choice under risk, complexity is often characterised by the number of payoff outcomes in a particular lottery (see for example Sonsino et al. 2002, Moffatt et al. 2015, Zilker et al. 2020, Fudenberg and Puri 2022, Puri 2024). Early research on this topic has found that *complexity aversion* is a common attribute in subjects' behaviour, that is they reveal a strong preference for simple lotteries over complex ones (lotteries with higher number of outcomes). Huck and Weizsäcker (1999) and Sonsino et al. (2002) were among the first to provide evidence that individuals discriminate heavily against complicated lotteries, such that even when the expected value was fixed, they still prefer the lotteries with fewer outcomes even when these lotteries have a higher variance. On the other hand, Wakker (2023, section 6, p.4) provides an extensive literature of studies that find either complexity seeking (see for example Birnbaum 2005 or Erev et al. 2017), or as much aversion as seeking (see among others Birnbaum et al. 2017 or Schmidt and Seidl 2014). Moffatt et al. (2015) estimate the distribution of attitudes towards complexity, finding that 50% are complexity-averse, 33% complexity-neutral, and only 17% complexity-loving. They also find that this rate of responsiveness to complexity reduces with experience to the extent that the average subject becomes almost complexity neutral by the end of the experiment. This convergence to complexity neutrality does not necessarily imply that subjects no longer have a distaste for complex tasks, instead it could suggest that they merely adopted a different strategy to make their decision, one which meant the complexity of the task was no longer hindering their decision process (i.e. heuristics).

From a theoretical modelling point of view, various expectation-based utility models (e.g. mean-variance, Expected Utility, Cumulative Prospect Theory) have been modified to capture complexity aversion. Moffatt et al. (2015) test versions of the mean-variance model, and expected utility, while Puri (2024) introduces Simplicity Theory which models decisions as expected utility less a cost for the complexity of the lottery, with several layers of generalization. Fudenberg and Puri (2022), propose a model based on the latter, that combines the standard cumulative prospect theory (CPT) model with a complexity cost. This model captures preferences for lotteries with smaller number of outcomes and they show that both probability weighting and complexity costs have an important role to play in predicting these risky alternatives. Diecidue et al. (2015) find that their results are consistent with prospect theory, but can also be explained by a population with heterogeneous aspiration levels. On the other hand, Bernheim and Sprenger (2020) find that PT and CPT fail rigorous tests that they design, and conclude that there is a possibility that the observed behaviour reflects a combination of standard CPT and a form of complexity aversion linked to heuristics, while Georgalos and Nabil (2023) show that the descriptive capacity of CPT is decreasing on the level of complexity in a dataset.

Previous research has also suggested that when decisions are more complex, individuals may avoid making a decision altogether, they might procrastinate, but more often than not they decide to stick with a default option or strategy (Iyengar and Lepper 2001; Thaler and Sunstein 2009). This led a strand of the literature to associate the distaste for complexity with an increase in one's cognitive load and therefore, an increase in reliance on simplified strategies or heuristics (rules of thumb). Venkatraman et al. (2014) shows that when faced with multiple-outcome gambles involving probabilities of both gains and losses, people often use simple heuristics that maximise the overall probability of winning. Coricelli et al. (2018) find that subjects may employ both a simplifying strategy and a compensatory strategy, providing evidence in support of a multiple-strategy approach to decision making. Oberholzer et al. (2024) report evidence of complexity aversion, suggesting a tendency to avoid cognitive effort as a potential explanation. Zeisberger (2022) suggest that the more complex the decision problem, the more likely it is the decision-maker will apply heuristics. Further studies have also supported the idea that complexity induces the use of heuristics with a focus on gain and loss probabilities (Erev et al. 2010; Payne 2005).

While this literature hints towards the extensive use of heuristics and simplification strategies as a response to the increased cognitive load that complexity causes, the relationship between heuristic decision making and complexity has not been thoroughly investigated. With this study, we aim to contribute to a recent literature that explores the impact of complexity on decision making. Recent work by Arrieta and Nielsen (2024) provides direct evidence that individuals adapt their decision-making processes in response to complexity by resorting to procedural strategies. They show that as decision environments become more complex (i.e. lotteries include more outcomes), individuals are more likely to adopt describable, step-by-step decision procedures. These procedural strategies simplify the decision process and reduce cognitive load, leading to more consistent choices under complexity. In our study, we take a different approach by examining complexity aversion and investigate how individuals facing complex lotteries tend to avoid the complexity altogether, preferring simpler lotteries and relying on a cognitive toolbox of heuristic strategies. Enke and Graeber (2023) introduce the concept of cognitive uncertainty, which represents individuals' subjective uncertainty over their own decision-making, especially in complex or uncertain environments. This uncertainty leads individuals to make decisions that are often simplified or "noisy", leading to systematic biases in economic decision-making, including probability weighting and conservatism in belief updating. Enke and Shubatt (2023) develop quantitative indices to measure both objective and subjective complexity in lottery choice problems, aiming to understand how increasing complexity leads to greater decision-making errors and biases. They show that as complexity increases, decision-making becomes noisier, leading to biases like spurious risk aversion due to the difficulty of accurately assessing complex lotteries. Finally, Georgalos and Nabil (2023) show that increased complexity, and therefore the increased cognitive fatigue, lead agents to resort to heuristic decision making (following simple rules of thumb) rather than using complicated expectation utility models.

In this paper we aim to explore the relationship between complexity aversion and heuristic decision making in the presence of extreme complexity. The heuristics literature assumes that people are equipped with a repertoire of heuristics (strategies) and simplifying processes (rules of thumb) to solve the tasks they face in daily life. This idea has been theoretically modelled with the aid of a *cognitive toolbox*, from which people might adaptively choose their respective strategies. Payne et al. (1993) argued that the decision makers are equipped with a set of strategies and select among them when faced with a decision; an approach which was later extended in Gigerenzer (2002) who models decision making as probabilistic draws from a toolbox of heuristic rules. Scheibehenne et al. (2013) propose a model of strategy selection. More specifically, they suggest a framework on how to quantitatively specify a toolbox model of cognition, and how to rigorously test it using Bayesian inference techniques. Using data from an experiment designed to elicit preferences towards risk and complexity aversion, we implement the methodology suggested in Scheibehenne et al. (2013) to estimate cognitive toolbox models. We then test these models against popular expectation-based utility models, modified to account for complexity aversion. We compare the models based on both their in-sample and out-of-sample (predictive) capacity. We find that for the majority of the subjects, a toolbox model of simple heuristics has better descriptive and prescriptive capacity than competing compensatory models.

## 2 Theoretical Framework

In this section we present the theoretical models designed to capture preferences towards complexity and risk. The subjective complexity of a choice task is generally characterised in the literature by the number of alternatives on the decision maker's choice set, or the number of payoff outcomes in a particular lottery (see among others Sonsino et al. 2002, Moffatt et al. 2015, Zilker et al. 2020, Fudenberg and Puri 2022, Puri 2024). In our comparison, we include three expectation-based utility models that have been developed or modified to account for this type of complexity, as well as a cognitive toolbox of heuristics. We include the two models tested in Moffatt et al. (2015), namely the *mean-variance* complexity aversion and the Viscusi (1989) *Prospective Reference Theory*, the *CPT-Simplicity Theory*, a recent Cumulative Prospect Theory specification to account for complexity, as proposed in Fudenberg and Puri (2022), and a toolbox model of simple heuristic rules, as proposed in Scheibehenne et al. (2013) and partially implemented in Stahl (2018).

### 2.1 Mean-Variance Complexity Aversion

This model assumes that the utility function of the decision maker takes into consideration the expected value of the lottery (mean), the variance (exposure to risk), and its complexity (measured by the number of outcomes)[1]. The utility function for an individual $i$ is given by:

$$U(p,x) = \mu_{(p,x)} - \alpha_i \sigma^2_{(p,x)} - \gamma_i C_{(p,x)} \tag{1}$$

where $\mu_{(p,x)}$ is the expected value of the J-outcome lottery $\mathcal{L} = \{p_1, x_1; \cdots ; p_J, x_J\}$ defined as:

$$\sum_{j=1}^{J} p_j x_j$$

[1]This corresponds to the type 4 model in Moffatt et al. (2015, p. 158) and consists of a modification of the mean-variance model to account for complexity aversion as a function of the number of outcomes. Similarly, this model can be also considered as special case of Puri (2024) Simplicity Theory.

$\sigma^2_{(p,x)}$ is the variance of the lottery defined as:

$$\sum_{j=1}^{J} p_j \left(x_j - \mu_{(p,x)}\right)^2$$

and $C_{(p,x)}$ is the measure of complexity of the lottery, operationalised as C=0 for a sure payoff, C=1 for a simple lottery, C=2 for a complex, and C=3 for a very complex lottery. The parameter $\alpha$ is closely related to the coefficient of absolute risk aversion, while $\gamma$ represents the degree of complexity aversion when $\gamma > 0$.

### 2.2 Prospective Reference Theory

This model assumes that the decision makers do not take the stated probabilities at face value, but act as Bayesians, and view the prior probability of each outcome of the lottery $\mathcal{L}$ as $1/J$. The model follows the same specification as above but replaces the objective probabilities in the expected value formula with transformed ones of the form:

$$\tilde{p}_j = \frac{\delta\frac{1}{J} + p_j}{\delta + 1},\ j = 1, \ldots, J; J > 1 \tag{2}$$

The parameter $\delta$ defines the degree of probability distortion. When $\delta \to 0$ the transformed probabilities coincide with the objective ones. On the contrary, as $\delta \to \infty$, $\tilde{p}_j \to 1/J$.

### 2.3 CPT-Simplicity Theory

CPT-Simplicity theory, introduced in Fudenberg and Puri (2022), modifies the CPT model to account for complexity aversion by introducing a complexity cost that captures a preference for lotteries with fewer number of outcomes. The CPT-simplicity model is defined as:

$$U(p,x) = \sum_{j=1}^{J} u(x_j) \left[ w\left(\sum_{k=1}^{j} p_k\right) - w\left(\sum_{k=1}^{j-1} p_k\right)\right] - C(|support(p)|) \tag{3}$$

where $C(x)$ is a three-parameter sigmoid cost function to account for complexity, specified as:

$$C(x) = \frac{\iota}{1 + e^{-\kappa(x-\rho)}} - \frac{\iota}{1 + e^{-\kappa(1-\rho)}} \tag{4}$$

with $x$ being the number of outcomes of a lottery, $\iota$ the height of the function, $\rho$ the midpoint of the rise, and $\kappa$ the slope, with larger values of $\kappa$ indicating a steeper slope[2]. The function satisfies the condition $C(1) = 0$, while $w(.)$ is the Tversky and Kahneman (1992) probability weighting function[3]:

$$w(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}} \tag{5}$$

Finally, a power (CRRA) utility function is assumed for the monetary payoff transformation.

### 2.4 Cognitive Toolbox

Following Scheibehenne et al. (2013), a toolbox model can be represented by a set of different psychological processes or strategies $f$, and each strategy predicts a particular course of action, depending on the *ecology* of the decision environment. The outcome of this process can be modelled with the aid of a mixture proportion parameter $\beta$, which indicates the probability of choosing each strategy in the toolbox, where $\beta$ a vector. For instance, for a particular toolbox TB consisting of $J$ strategies, each strategy $f_j$ will be selected with probability $\beta_j$, with $\sum_{j=1}^{J} \beta_j = 1$. For instance, a potential toolbox with 4 strategies would be defined as:

- Pick the lottery with the highest payoff (MAXIMAX) with probability $\beta_1$
- Pick the lottery with the highest minimum payoff (MINIMAX) with probability $\beta_2$
- Pick the lottery with the highest most likely payoff (MOST LIKELY) with probability $\beta_3$
- Calculate the arithmetic mean of all outcomes for each gamble and pick the lottery with the highest mean. (EQUI) with probability $1 - \sum_{i=1}^{3} \beta_i$

[2]Sigmoid functions have been extensively used in the artificial neural networks literature.

[3]We also tried different specifications of the probability weighting function (both one and two-parameter functionals) with TK being the best performing specification.

This modelling specification allows for the underlying cognitive process of strategy selection to remain unspecified, given that the value of the parameter vector $\beta$ will be estimated by the data, providing the empirical validation of the latent strategy mix. Given this mixture specification, the compound probability of choosing lottery $A$ can be specified based on the sum of the individual likelihoods of each $f_j$, weighted by the mixture probability $\beta_j$:

$$p(A|TB) = \sum_{j=1}^{J} [\beta_j \times P(A|f_j)] \tag{6}$$

where $P(A|f_j)$ is the individual predicted probability of each strategy. Since the most distinguishable feature of a toolbox model is its adaptive nature (each individual adopts their chosen strategies depending on the choice environment), we deviate from the standard practice of fixing a pre-determined set of strategies, and allow for heterogeneity between subjects, both in terms of size (how many strategies) and in terms of content (which strategies). The toolbox models we investigate can accommodate a variety of heuristics (out of a total of 10 heuristics extensively utilised in the literature[4]) and sizes (ranging from 2 to 5 strategies per toolbox[5]). We achieve so by estimating, subject-by-subject, every potential toolbox of size up to 5, that is formed as a combination of a subset of the available 10 heuristics. This gives in total 627 toolbox models.

## 3 Data

We re-analyse the data from Moffatt et al. (2015). This dataset involves 80 subjects participating in a 2-phase experiment, where in each phase subjects faced 27 tasks in which they were asked to choose between two lotteries with the same expected value, but with differing degrees of complexity and risk (phase 2 consisted of the same 27 tasks presented in a different order). The experiment was incentivised using the random lottery incentive mechanism. The experimental

[4]We use the heuristics studied in Glöckner and Pachur (2012). In the Online Appendix there is the full list of heuristics along with a description of the choice they prescribe.

[5]Scheibehenne et al. (2013) discuss how including too many strategies can lead to the strategy sprawl problem.

design builds on Sonsino et al. (2002) and Sitzia and Zizzo (2011) single period tasks. The construction of the lotteries is based on the two tasks presented below. The first task involves the choice between a sure win (SW) and a simple 3-outcome lottery ($S_3$).

$$SW = \begin{cases} 107, & \text{with probability } 1 \end{cases} \quad S_3 = \begin{cases} 80, & \text{with probability } 0.40 \\ 100, & \text{with probability } 0.30 \\ 150, & \text{with probability } 0.30 \end{cases}$$

Using the $S_3$ lottery and following a particular procedure [6], it is then possible to generate a *complex* lottery, with nine outcomes[7], and a *very complex* lottery with 27 outcomes. In vector form, this lottery can be written as $S_\alpha = (p, x) = \Big((p_1 p_2 p_3)', (x_1, x_2, x_3)'\Big)$. A complex lottery $C_\alpha$ can be generated from $S_\alpha$ using the formula:

$$C_\alpha = \left( vec(pp'); vec\left(\frac{1}{2}xi_3' + \frac{1}{2}i_3 x'\right)\right)$$

where $i_3$ is a vector of size 3 consisting of ones and $vec(A)$ is the function that transforms a $n \times n$ matrix $A$ into a $n^2 \times 1$ (column) vector consisting of the elements of $A$. This lottery is equivalent to playing $S_\alpha$ twice and using the arithmetic mean from the two plays as the outcome.

The new lottery will be more complex, but at the same time *safer*, since it will be characterised by a lower variance. On top of the SW lottery, they generated six simple, six complex and six very complex lotteries. Using three simple lotteries, they first generated three complex and three very complex lotteries. Then, using the so-called *safe* version of the simple lotteries, which has decreased spread of the extreme outcomes and an unchanged middle outcome, they constructed three further complex and three very complex *safe* lotteries. The pairwise combinations between a subset of these lotteries, along with the SW lottery, gives the total of the 27

[6] To save on space, we briefly describe the process in the Appendix and we refer the interested reader to the original study (Moffatt et al. 2015, p.151).

[7] Nevertheless, some of this outcomes will be identical and therefore, the complex lottery consists of only 6 outcomes.

tasks (see Moffatt et al. 2015, Table 2a, pp. 152-153 for the full set of tasks). All lotteries have the same expected value which also contributes to the complexity of the task.

## 4 Econometric Analysis and Results

We estimate all the models using Hierarchical Bayesian econometric techniques, which allow for the simultaneous estimation of individual level parameters and the hyper-parameters of the group level distributions (see Balcombe and Fraser 2015; Ferecatu and Önçüler 2016; Baillon et al. 2020; Alam et al. 2022 and Gao et al. 2022 for some recent applications of Bayesian econometrics in risky choice). We compare models both *in-sample*, and *out-of-sample*. Testing models both in-sample and out-of-sample is essential because in-sample testing evaluates how well the model fits the data it was trained on, while out-of-sample testing assesses the model's ability to generalise and predict unseen data, ensuring that the model is not overfitting and remains robust in real-world applications.

In all subsequent analyses, we fully leverage the flexibility of Bayesian methods to account for individual heterogeneity. All comparisons are based on estimates obtained at the subject level. Specifically, for both Toolbox and non-Toolbox models, we derive individual-level parameter estimates and use these to classify subjects into models for all subsequent analyses. In particular, we first compare the models in-sample, based on the value of the Bayes Factor, using the data from phase 1 of the experiment. We then compare the models based on their out-of-sample predictive capacity (predicted log-likelihood) on the phase 2 tasks, using the estimates from phase 1. To capture stochasticity in choice, we model the error structure assuming a logit link function. The probability of choosing lottery A is given by:

$$p(A,B) = \frac{\exp(\phi U_A(p,x))}{\exp(\phi U_A(p,x)) + \exp(\phi U_B(p,x))}$$

where $U(p,x)$ is the utility as defined in section 2, and $\phi$ an index of the sensitivity to differences in utility, to be estimated. The overall likelihood is a Bernoulli distribution that can be expressed as $P(D) = \prod p(A,B)^I \times (1 - p(A,B))^{(1-I)}$, where $I$ is an indicator function, taking the value 1 when the subject chose A, otherwise 0.

For the toolbox model, since the heuristics generate ordinal choice propensities (i.e. deterministic), we assume a *constant-error* choice rule to capture stochastic choice in the data, where the decision maker chooses with constant probability $1 - \varepsilon$, the option that the heuristic prescribes, and with probability $\varepsilon$ she makes a mistake[8]. The overall likelihood for a given subject is therefore the product, across all the tasks, of the weighted sum of predicted probabilities across the number of strategies in a given toolbox.

Table 1 reports the results of the classification. The first column classifies subjects to models based on the value of the Bayes Factor, while the second column, according to the models' predictive capacity. In-sample, the toolbox model has the best performance for 51.3% of the subjects, followed by the CPT-Simplicity Theory (22.5%), the mean variance complexity aversion (17.5%), and the Prospective Reference model which was best for 8.8% subjects. A similar pattern is also observed in our out-of-sample prediction exercise. The toolbox model is best for 70% of the subjects, followed by the mean variance complexity aversion (26.3%), the Prospective Reference Theory model (5%) and the CPT-Simplicity Theory model (3.8%).

In the results reported in the Table, we observe that, when considering confidence intervals, the performance difference between the best and the second-best model exceeds one standard error for all but two subjects. However, when applying a stricter criterion, requiring that the performance difference between the first and second-best models, between the second and third-best models, and so on, is greater than one standard error, we find that five subjects must be excluded from the analysis. Of these, three are classified as Toolbox decision makers, and two are classified as Mean-Variance Complexity Averse. Overall, for 14 subjects the Toolbox model performs poorly and is classified as last.

[8]This is the part $P(A|f_j)$ in Equation 6.

| Model | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Toolbox | 41 | 56 | 62 | 61 | 65 |
| % | 0.513 | 0.700 | 0.775 | 0.763 | 0.813 |
| Mean-variance complexity aversion | 14 | 21 | 15 | 1 | 7 |
| % | 0.175 | 0.263 | 0.188 | 0.013 | 0.088 |
| Prospective Reference Theory | 7 | 4 | 1 | 10 | 0 |
| % | 0.088 | 0.050 | 0.013 | 0.125 | 0.000 |
| CPT-Simplicity Theory | 18 | 3 | 2 | 8 | 8 |
| % | 0.225 | 0.038 | 0.025 | 0.100 | 0.100 |
| TOTAL | 80 | 80 | 80 | 80 | 80 |

Table 1: Number of subjects for which a model is classified as best, based on: (1) In-sample fit (Bayes Factor), (2) Out-of-sample fit (predicted log-likelihood on phase 2 data), (3) K-fold Cross Validation (mean predicted log-likelihood from 3 folds on phase 1 data), (4) In-sample fit (Bayes Factor on a subset of Moffatt et al. (2015) tasks with similar skewness), and; (5) Out-of-sample fit (predicted log-likelihood on phase 2 subset data using tasks with similar skewness).

| Model | In-sample | Out-of-sample | TLBX | MVCA | PRT | CPT-SIMPLE |
|---|---|---|---|---|---|---|
| TLBX | 41 | 56 | 31 | 7 | 3 | 0 |
| % | | | 0.756 | 0.171 | 0.073 | 0.000 |
| MVCA | 14 | 21 | 6 | 7 | 0 | 1 |
| % | | | 0.429 | 0.333 | 0.000 | 0.143 |
| PRT | 7 | 4 | 5 | 1 | 0 | 1 |
| % | | | 0.714 | 0.143 | 0.000 | 0.143 |
| CPT-SIMPLE | 18 | 3 | 10 | 6 | 1 | 1 |
| % | | | 0.556 | 0.333 | 0.056 | 0.056 |

Table 2: The Table shows the transition matrix of model classifications. The first two columns summarize the in-sample and out-of-sample classification of subjects. The remaining columns display the percentage of subjects from each in-sample classification who were reclassified into different models based on out-of-sample predictions. TLBX stands for Toolbox, MVCA for Mean-variance Complexity Aversion, PRT for Prospective Reference Theory and CPT-SIMPLE, for Simplicity Theory.

It is important to note that in-sample and out-of-sample model selection are distinct approaches that can yield different outcomes. In-sample selection identifies the model with the best descriptive fit to the observed data, while out-of-sample selection prioritizes predictive accuracy. Consequently, a subject classified as a Toolbox decision-maker in both contexts may not necessarily rely on the same set of tools. For example, a subject might be identified as a Toolbox decision-maker using four tools in-sample but classified as relying on only two tools—possibly different ones—out-of-sample. As a result, one may wonder how stable is this classification between in-sample and out-of-sample. Table 2 presents a transition matrix of types. The first two columns provide the same information as Table 1, detailing the classification of subjects across different models in and out-of-sample. The remaining columns show, within each model, how many subjects are classified as each model type based on out-of-sample predictive performance. For example, in the first row, of the 41 subjects classified as toolbox decision makers in-sample, 75.6% remain classified as toolbox decision makers out-of-sample, 17.1% are classified as Mean-variance Complexity Averse, 7.3% as Prospective Reference Theory decision makers, and none are classified according to CPT-Simplicity Theory. This matrix provides a clearer picture of the stability of model classifications and allows for a deeper understanding of how subjects' classifications change between in-sample and out-of-sample predictions. Furthermore, we investigated whether there was a pattern in the types of heuristics used by subjects who were classified as Toolbox decision makers in-sample but were classified differently out-of-sample. Among the seven subjects classified as Mean Variance Complexity Averse, six included Maximax (MAXI) in their toolbox, while Minimax (MINI) and Priority (PRIO) appeared in five out of seven. For the three subjects classified as Prospective Reference Theory, both Equal Weight (EQW) and Better Than Average (BTA) heuristics were present in all their toolboxes.

Given the performance of the toolbox model, we next focus on the size and the content of each toolbox. Figures 1 and 2 illustrate the distribution of the different sized toolboxes both in and out-of-sample. In both cases the majority of subjects' (who is classified as toolbox decision makers) use 4 or 5 heuristics, while very few use only 2. This outcome is in line with previous results in the literature (see Makridakis and Winkler 1983; Ashton and Ashton 1985; He et al. 2022). There seems to be a slight drop in the size of toolboxes, out-of-sample, which could be the effect of learning and increased familiarity with the task.

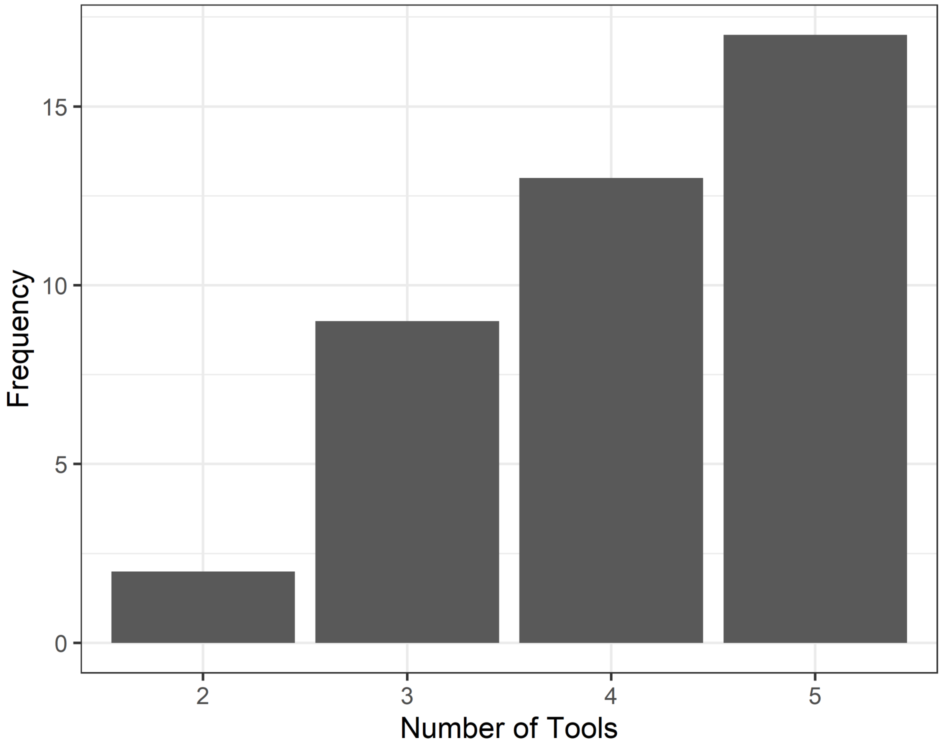


Figure 1: Frequency of toolbox sizes (in-sample).

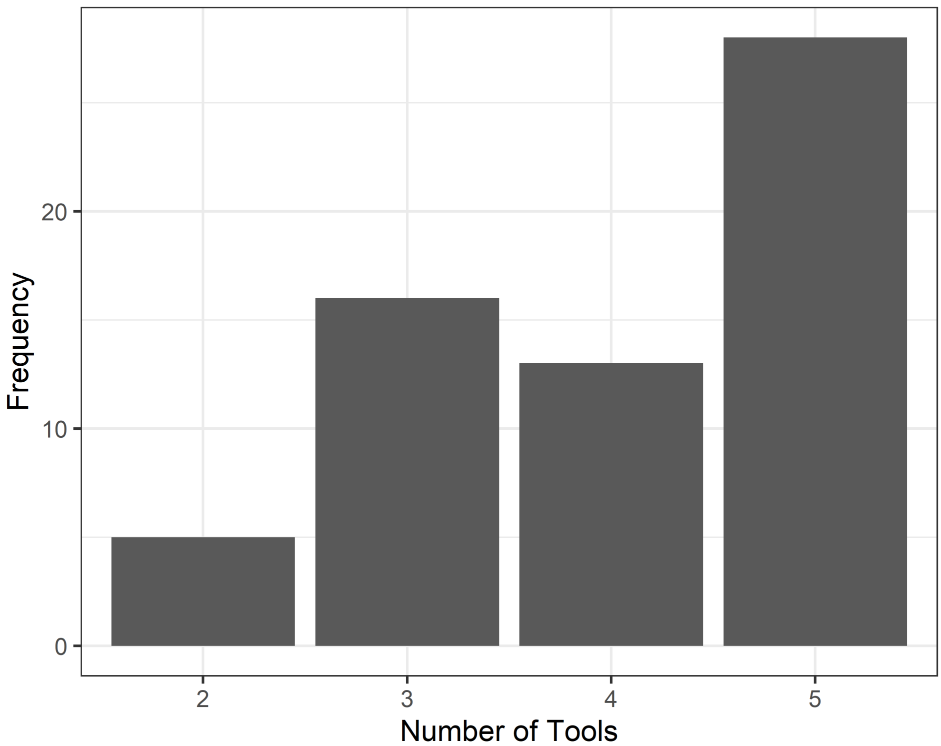


Figure 2: Frequency of toolbox sizes (out-of-sample).

Regarding the content of these toolboxes, Figures 3 and 4 illustrate the distribution of heuristics across all toolboxes, in and out-of-sample, respectively. Three heuristics outperformed all others, both in and out-of-sample as they were present in the majority of the toolboxes, namely, the Minimax (MINI-choose the gamble with the highest minimum outcome), the Least Likely (LL-identify each gamble's worst outcome and choose the gamble with the lowest probability of the worst outcome)and the Priority heuristic (PRIO-go through a number of reasons). Similarly, the three worst performing heuristics, both in and out-of-sample were the Lexicographic (LEX-go through a number of pairwise comparisons on the outcomes), the Equiprobable (EQUI-calculate the arithmetic mean of all outcomes for each gamble and choose the gamble with the highest mean) and the Most Likely (ML-identify each gamble's most likely outcome and choose the gamble with the highest, most likely outcome). Given the nature of these heuristics, it is easy to infer that subjects tend to resort to strategies that will protect them from the worst case scenario (i.e. worst outcome) or by guaranteeing a minimum payoff, while avoiding strategies that would expose them to higher levels of complexity due to the requirement of demanding calculations (e.g. calculating the mean of a gamble). When we compare in and out-of-sample we find strong evidence in favour of the Priority Heuristic (PRIO), a heuristic that has received much attention in the literature because of its capacity to explain risky choice. The PRIO is a lexicographic strategy that requires several rounds of reason comparing payoffs and probabilities and is therefore more cognitively demanding compared to simpler heuristics. This may be a potential explanation of the drop in complexity averse and seeking subjects that Moffatt et al. (2015) find in the phase 2 data.

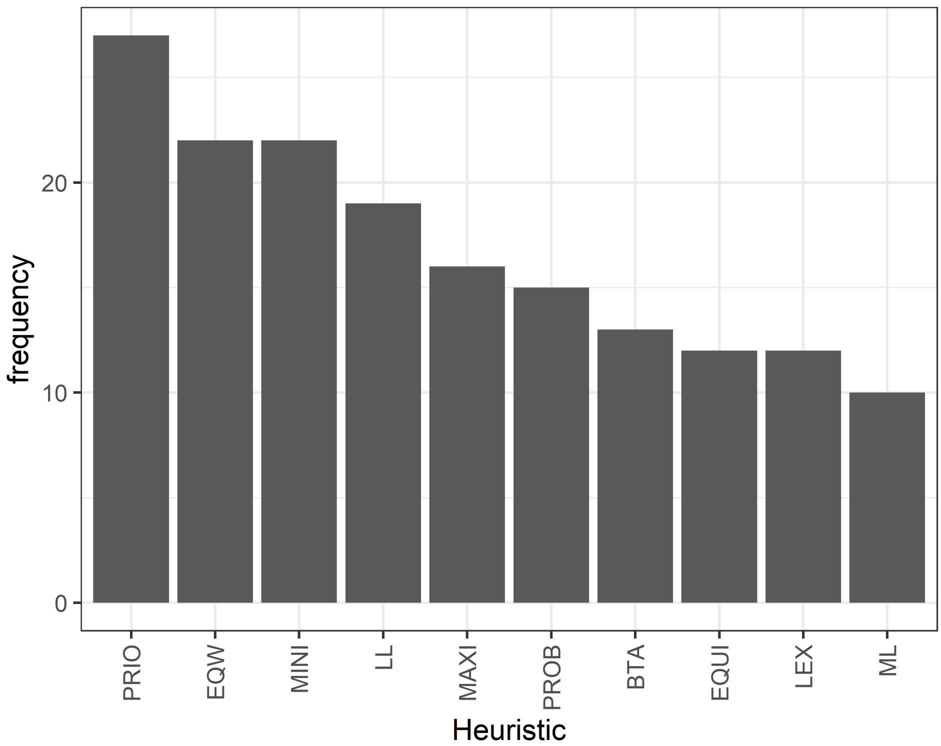


Figure 3: Frequency of heuristics (in-sample).

A potential issue with the out-of-sample prediction exercise, using data from phase 2, is that experience effects may drive the results, as subjects have already been exposed to the lotteries. As a robustness check, we re-estimate the models using the data from phase 1 and classifying subjects based on a K-fold cross validation exercise (James et al., 2013). In particular, we divided the tasks into 3 folds, with 9 tasks in each fold. For each iteration, we used the data from 2 folds to estimate the models (training set) and predicted the choices in the remaining fold (test set). This process was repeated 3 times, ensuring that each fold was used once as a test set and twice as part of the training set. Models were then classified based on their mean predictive performance across these iterations in this K-fold out-of-sample validation exercise. The results are reported in Table 1, column (3). The results are qualititavily similar to those reported above, with the vast majority of the subjects being classified as toolbox decision makers (77.5%).

A limitation in the experimental design in Moffatt (2016) is that it does not control for skewness, or in other words, the asymmetry in the distribution of outcomes. This could potentially lead to confounding effects where any observed behaviour might be influenced by skewness rather than just the number of outcomes (complexity). In a recent study Spiliopoulos and Hertwig (2023) show that as the number of outcomes increases, the influence of skewness on decision-making becomes more pronounced, finding that skewness-preference models were more prevalent in more complex environments. Subsequent studies that adopt the definition of complexity as the number of outcomes (Oberholzer et al. 2024; Puri 2024) control for skewness to ensure that any aversion to complexity could be attributed to the number of outcomes rather than the distribution. To control for differences in the skewness between lotteries, we repeat the same estimation using a sub-sample of the original data. In particular, using the data from phase 1, we exclude all the pairs of tasks that have one option offering a certain amount (9 tasks) and we also exclude all the pairs for which the difference in skewness between two options is greater than 0.15, as a proxy of same-skewness lotteries. This leaves a sub-sample of 14 pairs of lotteries for each subject[9]. We classify subjects based both on the in-sample fit and the out-of-sample predictive power from phase 2 data. Table 1 reports the results for both cases, column (4) and (5) respectively. Again, a similar pattern is observed regarding the number of subjects that are classified as toolbox decision makers.

[9] As there is no clear guidance on what constitutes a small difference, we opted for the 0.15 cut-off as it strikes a balance between maintaining a sufficient number of data points per subject to ensure robust parameter estimation and constraining the sample to lotteries with comparable skewness. For a cut-off value smaller than 0.05, only 5 lotteries satisfy this condition—a relatively low number given the complexity of the models. For cut-off values in the interval [0.05, 0.14), the number of lotteries increases to 9. Similarly, for cut-off values in the interval [0.15, 0.19), the number rises to 14, and for values greater or equal to 0.2, the number of lotteries reaches 18. We repeated the same exercise for a cut-off of 0.1 and 0.2 which confirm that the qualitative result remains unchanged, with the majority of the subjects being best described by a toolbox model.

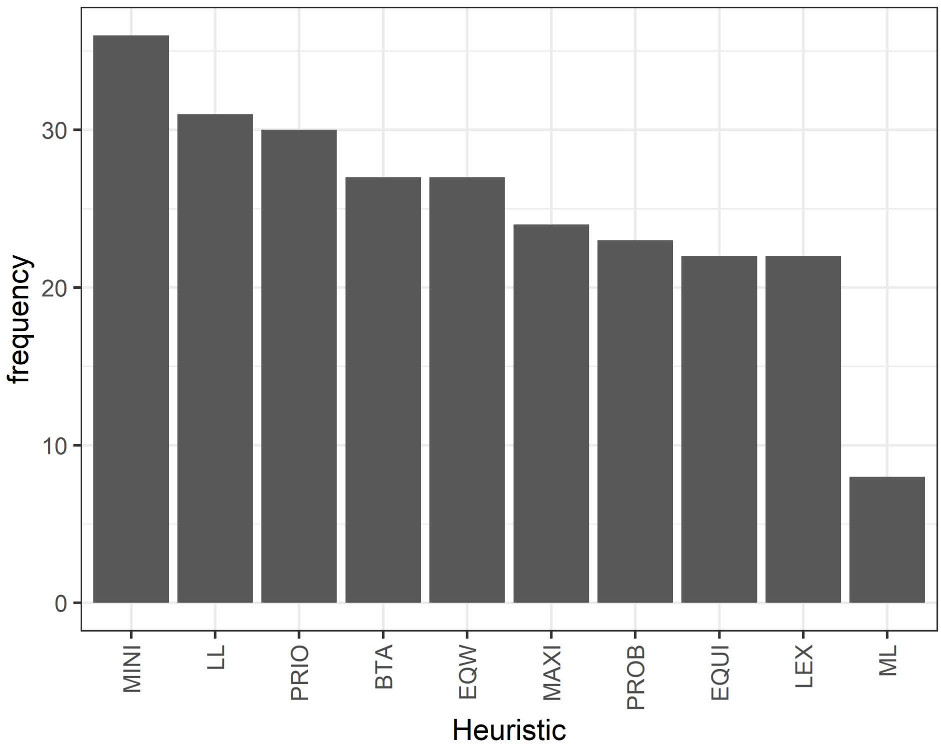


Figure 4: Frequency of heuristics (out-of-sample).

## 5 Conclusion

Our analysis highlights the importance of accounting for complexity when deciding on which explanatory model to adopt to describe individual behaviour. We have shown that with overly complex tasks comes increased cognitive fatigue in decision-making; a characteristic which heightens one's reliance on simple rules of thumb to make decisions. This results in an adaptive toolbox of heuristics outperforming other expectation-based models of decision-making, even when complexity aversion is captured within these competing parametric models. We provide a means of efficiently estimating structural models of decision-making, including a toolbox model, in-sample via the use of Bayesian Hierarchical modelling, and illustrate the robustness of these results in their alignment with our out-of-sample prediction results.

Ironically, analysing strategic processes and preferences in the face of increased complexity is a complex matter in itself, and it is easy to neglect vital attributes of complexity. Whilst most studies use the number of alternatives in a choice set as the key metric, we would urge future research to consider the works of Diecidue et al. (2015), Huck and Weizsäcker (1999) and Georgalos and Nabil (2023) who discuss how the formatting of probabilities and outcomes, the distribution moments (e.g. variance and mean) and other factors may well fall into the complexity function. The latter design a metric as a benchmark to determine a data sets complexity levels.

Finally we would urge future studies to expand decision-tasks beyond binary lotteries, as research has suggested the impact of complexity on risk taking is largely dependent on the decision format (Oberholzer et al., 2024). Before we jump to conclusions on complexity's effect on risky decision-making, we must ensure that numerous tasks of varying contexts and characteristics are examined, as it may be that the nature of certain tasks lead people to specific solutions.

## References

Alam, J., Georgalos, K., and Rolls, H. (2022). Risk Preferences, Gender Effects and Bayesian Econometrics. *Journal of Economic Behavior & Organization*, 202:168–183.

Arrieta, G. and Nielsen, K. (2024). Procedural decision-making in the face of complexity. *Working Paper*. Forthcoming.

Ashton, A. and Ashton, R. (1985). Aggregating Subjective Forecasts: Some Empirical Results. *Management Science*, 31(12):1499–1508.

Baillon, A., Bleichrodt, H., and Spinu, V. (2020). Searching for the Reference Point. *Management Science*, 66(1):93–112.

Balcombe, K. and Fraser, I. (2015). Parametric Preference Functionals under Risk in the Gain Domain: A Bayesian Analysis. *Journal of Risk and Uncertainty*, 50(2):161–187.

Bernheim, B. D. and Sprenger, C. (2020). On the empirical validity of cumulative prospect theory: Experimental evidence of rank-independent probability weighting. *Econometrica*, 88(4):1363–1409.

Birnbaum, M. H. (2005). Three new tests of independence that differentiate models of risky decision making. *Management Science*, 51(9):1346–1358.

Birnbaum, M. H., Schmidt, U., and Schneider, M. D. (2017). Testing independence conditions in the presence of errors and splitting effects. *Journal of Risk and Uncertainty*, 54(3):269–289.

Brandstätter, E., Gigerenzer, G., and Hertwig, R. (2006). The Priority Heuristic: Making Choices without Trade-Offs. *Psychological Review*, 113:409–432.

Coricelli, G., Diecidue, E., and Zaffuto, F. D. (2018). Evidence for multiple strategies in choice under risk. *Journal of Risk and Uncertainty*, 56(3):193–210.

Diecidue, E., Levy, M., and van de Ven, J. (2015). No aspiration to win? an experimental test of the aspiration level model. *Journal of Risk and Uncertainty*, 51:245–266.

Enke, B. and Graeber, T. (2023). Cognitive uncertainty. *The Quarterly Journal of Economics*, 138(4):2021–2067. Advance Access publication on May 27, 2023.

Enke, B. and Shubatt, C. (2023). Quantifying lottery choice complexity. *SSRN Electronic Journal*.

Erev, I., Ert, E., Plonsky, O., Cohen, D., and Cohen, O. (2017). From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychological Review*, 124(4):369–409.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., Hertwig, R., Stewart, T., West, R., and Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, 23(1):15–47.

Ferecatu, A. and Önçüler, A. (2016). Heterogeneous Risk and Time Preferences. *Journal of Risk and Uncertainty*, 53(1):1–28.

Fudenberg, D. and Puri, I. (2022). Simplicity and Probability Weighting in Choice under Risk. *American Economic Review Papers & Proceedings*, 112:421–425.

Gao, X., Harrison, G., and Tchernis, R. (2022). Behavioral Welfare Economics and Risk Preferences: a Bayesian Approach. *Experimental Economics*.

Georgalos, K. and Nabil, N. (2023). Testing models of complexity aversion. Economics Working Paper Series 2023/008, Lancaster University Management School.

Gigerenzer, G. (2002). The Adaptive Toolbox. In Gigerenzer, G. and Selten, R., editors, *Bounded Rationality: The Adpative Toolbox*, pages 37–50. Cambridge, MA: MIT Press.

Glöckner, A. and Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, (123):21–32.

He, L., Analytis, P., and Bhatia, S. (2022). The Wisdom of Model Crowds. *Management Science*, 68(5):3175–3973.

Huck, S. and Weizsäcker, G. (1999). Risk, complexity, and deviations from expected-value maximization: Results of a lottery choice experiment. *Journal of Economic Psychology*, 20:699–715.

Iyengar, S. and Lepper, M. (2001). When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology*, 79:995–1006.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer Texts in Statistics. Springer, New York, 1st edition.

Makridakis, S. and Winkler, R. (1983). Averages of Forecasts: Some Empirical Results. *Management Science*, 29(9):987–996.

Moffatt, P. (2016). *Experimetrics*. Macmillan Palgrave.

Moffatt, P., Sitzia, S., and Zizzo, D. (2015). Heterogeneity in Preferences Towards Complexity. *Journal of Risk and Uncertainty*, 51:147–170.

Oberholzer, Y., Olschewski, S., and Scheibehenne, B. (2024). Complexity aversion in risky choices and valuations: Moderators and possible causes. *Journal of Economic Psychology*, 100:102681.

Payne, J., Bettman, J., and Johnson, E. (1993). *The Adaptive Decision Maker*. Cambridge University Press.

Payne, J. W. (2005). It is whether you win or lose: The importance of the overall probabilities of winning or losing in risky choice. *Journal of Risk and Uncertainty*, 30(1):5–19.

Puri, I. (2024). Simplicity and risk. *Journal of Finance*, page 93. Forthcoming, posted 24 June 2024.

Scheibehenne, B., Rieskamp, J., and E, W. (2013). Testing Adaptive Toolbox Models: A Bayesian Hierarchical Approach. *Phychological Review*, 120(1):39–64.

Schmidt, U. and Seidl, C. (2014). Reconsidering the common ratio effect: The roles of compound independence, reduction, and coalescing. *Theory and Decision*, 77(2):267–284.

Sitzia, S. and Zizzo, D. J. (2011). Does product complexity matter for competition in experimental retail markets? *Theory and Decision*, 70(1):65–82.

Sonsino, D., Benzion, U., and Mador, G. (2002). The Complexity Effects on Choice under Uncertainty. *The Economic Journal*, 112(482):9.

Spiliopoulos, L. and Hertwig, R. (2023). Variance, skewness, and multiple outcomes in described and experienced prospects: Can one descriptive model capture it all? *Journal of Experimental Psychology: General*, 152(4):1188–1222.

Stahl, D. (2018). An Empirical Evaluation of the Toolbox Model of Lottery Choices. *The Review of Economics and Statistics*, 100(3):528–534.

Thaler, R. H. and Sunstein, C. R. (2009). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Penguin Books.

Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science*, 25:219–225.

Tversky, A. and Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Venkatraman, V., Payne, J. W., and Huettel, S. A. (2014). An overall probability of winning heuristic for complex risky decisions: Choice and eye fixation evidence. *Organizational Behavior and Human Decision Processes*, 125(2):73–87.

Viscusi, W. K. (1989). Prospective reference theory: Toward an explanation of the paradoxes. *Journal of Risk and Uncertainty*, 2(3):235–263.

Wakker, P. P. (2023). A criticism of bernheim and sprenger's (2020) tests of rank dependence. *Journal of Behavioral and Experimental Economics*, 107:102536.

Zeisberger, S. (2022). Do people care about loss probabilities? *Journal of Risk and Uncertainty*, 65(2):185–213.

Zilker, V., Hertwig, R., and Pachur, T. (2020). Age differences in risk attitude are shaped by option complexity. *Journal of Experimental Psychology: General*, 149(9):1644–1683.

# Appendix A Tasks

This Appendix briefly describes the procedure Moffatt et al. (2015) are using to generate the lotteries for their experiment. Consider the following lottery $S_\alpha$:

$$S_3 = \begin{cases} 80, & \text{with probability } 0.40 \\ 100, & \text{with probability } 0.30 \\ 150, & \text{with probability } 0.30 \end{cases}$$

This simple lottery with 3 outcomes, can generate a complex lottery with 9 outcomes, and a very complex lottery with 27 outcomes. In vector form, this lottery can be written as $S_\alpha = (p, x) = \Big((p_1 p_2 p_3)', (x_1, x_2, x_3)'\Big)$. A complex lottery $C_\alpha$ can be generated from $S_\alpha$ using the formula:

$$C_\alpha = \left(vec(pp'); vec\left(\frac{1}{2}xi_3' + \frac{1}{2}i_3 x'\right)\right)$$

where $i_3$ is a vector of size 3 consisting of ones and $vec(A)$ is the function that transforms a $n \times n$ matrix $A$ into a $n^2 \times 1$ (column) vector consisting of the elements of $A$. This lottery is equivalent to playing $S_\alpha$ twice and using the arithmetic mean outcome from the two plays as the outcome.

Applying this to the above lottery, we get:

$$vec(p \times p') = \begin{bmatrix} 0.16 & 0.12 & 0.12 \\ 0.12 & 0.09 & 0.09 \\ 0.12 & 0.09 & 0.09 \end{bmatrix} = \begin{bmatrix} 0.16 & 0.12 & 0.12 & 0.12 & 0.09 & 0.09 & 0.12 & 0.09 & 0.09 \end{bmatrix}$$

and $vec(pp'$ generates a vector of size nine with the element of the $p \times p'$ matrix. Then, for the payoffs:

$$vec\left(\frac{1}{2}xi_3' + \frac{1}{2}i_3 x'\right) = \begin{bmatrix} 80 & 90 & 115 & 90 & 100 & 125 & 115 & 125 & 150 \end{bmatrix}$$

which gives the lottery

$$C_3 = \begin{cases} 80, & \text{with probability } 0.16 \\ 90, & \text{with probability } 0.24 \\ 100, & \text{with probability } 0.09 \\ 115, & \text{with probability } 0.24 \\ 125, & \text{with probability } 0.18 \\ 150, & \text{with probability } 0.09 \end{cases} \quad S_3 = \begin{cases} 80, & \text{with probability } 0.40 \\ 100, & \text{with probability } 0.30 \\ 150, & \text{with probability } 0.30 \end{cases}$$

Using a similar procedure, it is possible to create a very complex lottery with 27 outcomes. For the full set of tasks please see Moffatt et al. (2015, Table 2a, pp. 152-153).

## Appendix B List of Heuristics

Table 3: Table of heuristics

| | Heuristic | Description |
|---|---|---|
| 1. | Priority Heuristic (PRIO) | Go through reasons in the order of: minimum gain, probability of minimum gain, and maximum gain. Stop examination if the minimum gains differs by 1/10 (or more) of the maximum gain; otherwise, stop examination if probabilities differ by 1/10 (or more) of the probability scale. Choose the gamble with the more attractive gain (probability). |
| 2. | Equiprobable (EQUI) | Calculate the arithmetic mean of all outcomes for each gamble. Choose the gamble with the highest mean. |
| 3. | Equal-weight (EQW) | Calculate the sum of all outcomes for each gamble. Choose the gamble with the highest sum. |
| 4. | Better than average (BTA) | Calculate the grand average of all outcomes from all gambles. For each gamble, count the number of outcomes equal to or above the grand average. Then choose the gamble with the highest number of such outcomes. |
| 5. | Probable (PROB) | Categorize probabilities as probable (i.e., $\geq 1/2$ for a two-outcome gamble, $\geq 1/3$ for a three-outcome gamble, etc.) or improbable. Cancel improbable outcomes. Then calculate the arithmetic mean of the probable outcomes for each gamble. Finally, choose the gamble with the highest mean. |
| 6. | Minimax (MINI) | Choose the gamble with highest minimum outcome. |
| 7. | Maximax (MAXI) | Choose the gamble with the highest outcome. |
| 8. | Lexicographic (LEX) | Determine the most likely outcome of each gamble and choose the gamble with the better outcome. If both outcomes are equal, determine the second most likely outcome of each gamble, and choose the gamble with the better (second most likely) outcome. Proceed until a decision is reached. |
| 9. | Least likely (LL) | Identify each gamble's worst outcome. Then choose the gamble with the lowest probability of the worst outcome. |
| 10. | Most likely (ML) | Identify each gamble's most likely outcome. Then choose the gamble with the highest, most likely outcome. |

Heuristics are from Thorngate (1980) and Payne et al. (1993), later used in Brandstätter et al. (2006) and Glöckner and Pachur (2012).