Economics Working Paper Series

2023/010

Heuristics Unveiled: A Comparative Analysis of Toolbox Models and Prospect Theory in Risky Choice

Konstantinos Georgalos and Nathan Nabil

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

LUMS home page: http://www.lancaster.ac.uk/lums/

Heuristics Unveiled: A Comparative Analysis of Toolbox Models and Prospect Theory in Risky Choice

Konstantinos Georgalos[*] Nathan Nabil[†]

October 31, 2023

Abstract

In an attempt to elucidate the classic violations of expected utility theory, the behavioural economics literature heavily relies on the influential work of Tversky and Kahneman (1992), specifically the Cumulative Prospect Theory (CPT) model and the Heuristics-and-Biases program. While both approaches have significantly contributed to our understanding of decision-making under uncertainty, empirical evidence remains inconclusive. In this study, we investigate the performance of each approach across a wide range of choice environments and increasing cognitive load, encompassing gains, losses, time pressure, and complexity. Utilising data from various studies and employing Bayesian inference, we assess the performance of CPT in comparison to an adaptive cognitive toolbox model of heuristics. For subjects classified as toolbox decision makers, we examine the content (i.e., which heuristics) and the size of the toolbox (i.e., how many heuristics). Our findings reveal that as the choice environment objectively increases in complexity, individuals transition from using sophisticated expectation-based utility models to relying on a set of simplification heuristics for decision-making. We quantify the relationship between toolbox usage and complexity, showing a significant and positive correlation between the two. Furthermore, our results indicate that as task complexity rises, individuals tend to employ smaller toolboxes with fewer heuristics for decision-making.

Keywords: Complexity · 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,
[†]Department of Economics, Lancaster University Management School, LA1 4YX, Lancaster, U.K.
✉ n.nabil@lancaster.ac.uk
We are grateful to Han Bleichrodt, Adam Sanjurjo, participants to the 2022 ESA European Meeting for providing useful comments, and to all the authors for making their data available. The support of The High End Computing (HEC) facility at Lancaster University is gratefully recognised.

1 Introduction

Underlying much of the neoclassical economic theory are a few key assumptions regarding agent's behaviour, successfully encapsulated in the notion of Bayesian rationality. This includes the presence of well-defined preferences, the ability of the agents to optimise behaviour in static and dynamic problems, the capacity of the decision makers to use probabilities to describe risky and ambiguous situations, and also to update these probabilities according to Bayes rule, upon the arrival of new information. Extensive empirical evidence from experiments in economics and psychology have consistently shown that individuals occasionally deviate from the predictions of the Bayesian paradigm. In an effort to understand these deviations from rationality and provide explanations to classic violations of expected utility theory, the behavioural economics literature largely depends on the influential work of Daniel Kahneman, Amos Tversky, and other researchers, namely the *Cumulative Prospect Theory* model and the *Heuristics-and-Biases* program. As Pachur et al. (2017) note, these constitute the two most commonly used approaches to model risky choice, the algebraic, expectation-based, compensatory models on the one hand (Hilbert 2011; Kahneman and Tversky 1979; Zindel et al. 2014) and models of simple rules of thumb (heuristics) on the other (Tversky and Kahneman 1974; Gigerenzer and Gaissmaier 2011). The former predicts overt decisions, combining components of risk attitudes such as utility curvature, probability weighting and loss aversion, whilst the latter, is concerned with capturing the underlying cognitive process. These two approaches radically differ in their assumptions and the way they model decision making, and both have been extensively used in the literature, to explain a number of paradoxes, such as the Allais paradox, the four-fold pattern, the certainty effect, the possibility effect, and intransitivities.

We aim to explore how individuals utilise either compensatory models (e.g. CPT), or simple rules of thumb (heuristics) in various decision-making domains that may increase ones cognitive load, specifically in the gain domain, the loss domain, under time pressure, and when subject to overly complex tasks.

A stream of evidence has proven CPTs descriptive power in the *gain* domain (Stott 2006; Fennema and Wakker 1997; Hey and Orme 1994; Loomes et al. 2002) supporting its robust explanation as to why we deviate from the axioms of expected utility. Whilst on the other hand, many have postulated that CPT's flexibility overpowers its descriptive capacity in some situations, and that individuals, dependent on the specific environment/context of choice, do not always engage in such complex subconscious algorithms when making decisions. It has been suggested that it is, instead, general rules of thumb that dominate (Brandstätter et al.; 2006).

Similarly, the *loss* domain has different psychological implications to the gain domain, so individuals are likely to have different preferences and strategies as the outcomes of a loss task will have different behavioural effects on the individual. Losses may increase an individual's cognitive load due to the associated higher subjective weighting and the additional mental strain of enduring a potential loss (Tom et al. 2007; Thaler 2000; Gal and Rucker 2018). The increase in cognitive load may result in a strategy alteration due to impaired judgment. The heightened stress induced by the potential loss forces individuals to rely on simple heuristics as a means of simplifying the task at hand and reducing potential losses (Payne; 1976). In the loss domain, Zeisberger (2022) find that decision-makers, in fact, pay explicit attention to loss probabilities in a loss context, which is completely irrespective of loss size. This result is supported by, implicit, experimental evidence on individuals providing a focus on loss probabilities in lottery tasks (Payne et al. 1980; Lopes and Oden 1999; Sokolowska 2006).

The existing literature of choice under *time pressure* suggests that individuals, due to a lack of time and commitment, will rely on short-cuts when making choices (del Campo et al. 2016; Payne et al. 1993). This increase in an individual's cognitive load may facilitate that switch from the thorough analysis of a decision based on cognition (the use of a compensatory model) to making a choice based on a simple heuristic. Payne et al. (1988) showed that heuristics, in particular the lexicographic rule, were more accurate than normative procedures (e.g. additive utility) when individuals had severe time deadlines. Similarly, Payne et al. (1996) showed that opportunity cost time pressure meant the subjects in their study adapted their decision-making procedure by processing less information, being more selective in their processing, and thus processing more by attribute; all characteristics that provide support for the use of heuristics in decision-making. Saqib and Chan (2015) point out that, under time pressure, individuals see the maximal possible outcome (e.g., best gain in gain domain and worst loss in loss domain) as more likely to happen, while Olschewski and Rieskamp (2021) find that risk preferences or strategy selection failed to explain the changes in behaviour due to time pressure. Finally, the question of how individuals will devise a strategy, make decisions, or alter preferences when presented with tasks of a higher *complexity* has been touched upon in the literature. Sonsino et al. (2002) find that individuals discriminate heavily against complicated lotteries, such that even when the expected value was fixed, and this information was disclosed to participants, they still prefer the lotteries with fewer outcomes even when these lotteries have a higher variance. Moffatt et al. (2015) find 67% of their subjects respond to complexity, and the majority of these display moderate levels of complexity aversion, while (Oberholzer et al.; 2021) study how this complexity may prompt the development of new strategies that make the task at hand less overwhelming. On the CPT side of the argument, Fudenberg and Puri (2022), conclude that probability weighting has an important role to play in predicting risky alternatives, while Bernheim and Sprenger (2020) find that PT and CPT fail rigorous tests that they design. Zeisberger (2022) suggest that the more complex the decision problem, the more likely it is the decision-maker will resort to heuristics such as lexicographic rules and potentially the "least likely" heuristic. 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).

Interestingly, these ideas are implicitly supported by existing evidence in the eye-tracking literature. Multiple studies have used process data to investigate and characterise decision strategies (see for example Fiedler and Glöckner 2012; Venkatraman et al. 2014; Harrison and Swarthout 2019). Notably, Arieli et al. (2011) find that the eye patterns of individuals suggest that subjects compare prizes and probabilities separately (focus either on prizes or probabilities depending on the nature of the task), which opposes more holistic cognitive bias model approaches (e.g. EUT or CPT) and provides evidence in favour of heuristic-type behaviour. It was also found that this effect was more pronounced when the suggested weighting of probabilities and utilities are more laborious to compute, something we wish to look into further. Nonetheless, there are studies that find that evidence from the eye tracking data opposes certain heuristics (Glöckner and Herbold; 2011), and at the same time, these studies tend to also rule out the idea that individuals look to maximise any form of expectations models. All of the above evidence confirms that the debate between CPT and Heuristics is still very much alive and active.

Many theories of human cognition assume that people are equipped with a repertoire of heuristics and simplifying processes to solve the tasks they face in daily life. In the literature, this idea has been theoretically modelled with the aid of a cognitive toolbox, from which people might adaptively choose their respective strategies. In the field of judgement and decision making, this concept was pioneered by Payne et al. (1993) arguing 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), modelling decision making as probabilistic draws from a toolbox of heuristic rules. While this modelling approach has been extensively investigated in the field of psychology and related studies in various domains, such as resource allocation, estimation and judgment of frequencies, skill acquisition, and learning processes, there is a notable absence of empirical evidence regarding the performance of these toolbox models of cognition, specifically in the context of risky choice. As Scheibehenne et al. (2013) highlight, while this theoretical framework of a cognitive toolbox provides a plausible account of intra- and interindividual differences in human behaviour, it is often unclear how to rigorously test the toolbox framework, how to quantitatively specify such a model, how to limit the number of toolbox strategies in the model to avoid the so called *strategy sprawl*, and how to formally test against alternative theories. To address these issues comprehensively, the authors propose the utilisation of Bayesian inference techniques. In this paper, we adopt the statistical framework proposed by Scheibehenne et al. (2013) to rigorously assess the performance of an adaptive toolbox model of heuristics in the field of risky choice compared to the benchmark CPT model. We employ this framework across a diverse range of potential heuristics, testing the model's performance in various domains and environmental contexts. Specifically, we aim to address the following research questions:

Research Question 1. *Can a toolbox model of simple heuristics explain lottery choices better relative to a sophisticated compensatory utility model (i.e. CPT)?*

Recent literature has primarily focused on one-to-one comparisons between a single heuristic and a flexible model with free parameters (most commonly CPT), providing overwhelming support in favour of the latter (Brandstätter et al. 2006; Rieger et al. 2008; Glöckner and Pachur 2012; Balcombe and Fraser 2015; Peterson et al. 2021). Nevertheless, this approach ignores the concept of *ecological rationality*, that is the fit between a heuristic or decision strategy and a choice environment which gives the agent the flexibility to adapt her strategy according to the decision task at hand. Various heuristics have been shown to take place simultaneously, so to better understand the behaviour of economic agents, it is imperative to determine which heuristics are used in which circumstances (del Campo et al.; 2016).

In the field of risky choice, there is a notable scarcity of research examining the performance of toolbox models. Stahl (2018) investigates whether a toolbox model of simple heuristic rules can help explain choices under risk relative to expected utility theory (EUT). The study concludes that *[· · ·] if we want to forecast the future lottery choices of humans and have limited prior data on which to make those forecasts, then our analysis suggests it would be better to use the Expected Utility Theory-Only model even if we believe it is not the true data generating process rather than using an overfitted toolbox model*. Mohnert et al. (2019) develop a model according to which the decision maker selects a decision strategy for a given choice problem rationally from a toolbox of strategies and they estimate the content of the toolbox at the individual level. Their adaptive toolbox model predicted people's risky choices better than single strategies, non-adaptive toolbox models, or adaptive toolbox models that assume that all decision , but performed worse than CPT. Finally, Olschewski and Rieskamp (2021) explore whether time pressure motivates subjects to use simple, noncompensatory strategies in a risky choice experiment. They assume

3 potential heuristics, and they find a slight but insignificant increase in the number of participants who resort to heuristics under time pressure, attributing this to increased noise in subjects' behaviour[1]. Nevertheless, prior efforts to assess toolbox models face obstacles stemming from simplification assumptions and econometric challenges in accurately characterising and identifying the model. These simplification issues arise in two primary forms. Firstly, with the exception of Mohnert et al. (2019), existing studies have not allowed for individual heterogeneity in toolboxes, implying that all subjects employ the same limited set of 3 or 4 heuristics. The second concern, raised by Stahl (2018), revolves around over-fitting induced by the estimation method. Notably, all the aforementioned studies employ Maximum Likelihood Estimation (MLE) techniques, which are known to produce estimates that are extreme, relatively noisy, and less reliable when compared to more flexible estimation methods, such as Bayesian Hierarchical modelling or Simulated Maximum Likelihood, that account for behavior at both individual and population levels. This leads us to our second question:

Research Question 2. *What is the best way to quantitatively specify and robustly estimate a toolbox model of cognition?*

Stahl (2018) using both simulated and actual data, suggests that it may be preferable to use the EUT-Only model, even when it is not believed to be the true data generating process, rather than employing an overfitted toolbox model. Building on this insight, we re-analyse the same datasets and investigate two potential explanations for this result (1) the choice of estimation method and (2) the nature of the data[2]. Regarding the former, Bishop (2006, pp. 166) cautions that "the use of maximum likelihood, or equivalently least squares, can lead to severe over-fitting if complex models are trained using data sets of limited size", and goes on to suggest "the phenomenon of over-fitting is really an unfortunate property of maximum likelihood and does not arise when we marginalize over parameters in a Bayesian setting". Using Bayesian inference, we demonstrate that the overfitting issue is significantly mitigated, resulting in more robust and reliable estimates. With regards to the second point, we explore how the nature of the dataset, and in particular the experimental design, can explain the poor forecasting performance of the toolbox model. Therefore, the next question we explore can be summarised as:

[1] A relevant study is He et al. (2022) focusing on the recent literature of collective model wisdom (model crowds) in decision analysis. It conducts a large scale comparison of 58 prominent models of risky choice and they find that crowds of risky choice models perform better than individual models. While they include a large number of popular heuristics in the candidate models, they do not explicitly test toolbox models of heuristics and their approach differs from ours in several aspects (i.e.type of lotteries, model selection, estimation method, CPT specifications, and number and type of heuristics).

[2] To save on space we delegate this replication to an Appendix.

Research Question 3. *What is the relationship between increased cognitive load and the utilisation of cognitive toolboxes?*

Stahl (2018) analyses the data of Hey and Orme (1994) and Harrison and Rutström (2008) which consist of 3-outcome binary lotteries, in the gain domain, with only four potential monetary outcomes and varying probabilities. We argue that this design and decision environment, may be more "user friendly" for expectation based models (i.e. Expected Utility). We extend our comparison of CPT and Toolbox models in various domains and conditions, specifically looking to explore whether decision-makers resort to the use of heuristics when the tasks at hand are more cognitively demanding. We therefore, explore the performance of the toolbox model, in a number of domains/environments that require more cognitive effort, including losses and mixed gambles, time pressure and complexity. Using features of a dataset such as the number of alternatives in a choice set, the formatting of probabilities and outcomes, or the distribution moments, we set up an index of complexity which enables us to quantify the relationship between increased cognitive load and the utilisation of cognitive toolboxes, and show how the two are correlated.

Given the flexibility that our econometric approach offers, we are able to estimate a large set of models per subject, and identify the combination of number and type of heuristics that best describe the behaviour of each individual subject. The next two questions focus on the size and the content of the toolbox model.

Research Question 4. *What is the optimal number of heuristic strategies that one should include in a toolbox?*

As mentioned before, the previous literature on toolbox modelling makes the simplification assumption that all the subjects share the same limited and fixed set of heuristics (3 or 4). We relax this hypothesis and allow the toolbox size to vary between 2 and 5 heuristics, accommodating individual-specific combinations of size and heuristic number. Furthermore, the toolboxes can comprise various combinations drawn from a comprehensive set of 11 heuristics extensively studied in the literature. Next, we inquire into the contents of these toolboxes, identifying which heuristics are most frequently employed and what insights they provide into the subjects' risk preferences.

Research Question 5. *What is inside the toolbox?*

After identifying all the subjects for whom the toolbox model is the best specification, we can not only assess the size of the toolboxes (i.e., the number of heuristics) but also delve into the content of these toolboxes. This allows us to determine which heuristics are most frequently employed in different environments and gain insights regarding the risk preferences of the subjects.

To summarise, we estimate toolbox models of cognition using data from prominent studies in various domains and contexts (i.e., gain, loss and mixed domains, time pressure, and complexity). These toolbox models comprise a comprehensive set of 11 heuristics, and we compare them with four different CPT model specifications. We identify the domains and environments in which decision makers resort to the use of simplifying strategies, with the percentage of subjects classified as Toolbox decision makers, ranging from 2 to 67%. We find that for the subjects classified as Toolbox decision makers, the majority is using 3 or 4 strategies. Furthermore, we observe extensive heterogeity regarding the type of heuristics subjects employ, with the vast majority resorting to heuristics that provide a safety net (e.g. Minimax or Least Likely), while risk seeking strategies are least preferred (e.g. Maximax). We introduce a metric to gauge the complexity of experimental stimuli, and based on this index, we observe a positive and significant correlation between complexity and the utilisation of heuristics in addressing risky choice problems. Finally, our analysis allows us to identify the domains/environments where CPT can explain behaviour better, which allows us to run a racehorse between different probability weighting functions, showing that a two-parameter weighting function always performs better compared to the one-parameter family of weighting functions.

The rest of the paper is organised as follows. Section 2 outlines the theoretical frameworks of the the two competing models, section 3 provides the details of our econometric approach, section 4 briefly presents the datasets we employ, while section 5 presents the results. We then conclude.

2 Theoretical Framework

In this section, we present the underlying assumptions of the two decision making models that we estimate. In our analysis, the value of the outcomes can be positive, zero or negative, depending on the context and the environment that we explore. We present the full specification (i.e. the one that accounts for both gains and losses) which we adapt accordingly. We present both the deterministic assumptions of each model, as well as the stochastic assumptions we make, in order to take into account noise and heterogeneity in decision making.

2.1 Cumulative Prospect Theory

The decision maker faces pairs of n-outcome lotteries with outcomes $x_1 \leq \ldots \leq x_k \leq 0 \leq x_{k+1} \leq \ldots \leq x_n$ and corresponding probabilities $p_1 \ldots p_n$. Following Tversky and Kahneman (1992) we assume that a decision maker is endowed with a utility function $u(.)$, on monetary outcomes with $u(0) = 0$, and a probability weighting function $w(p)$, that transforms the objective probabilities into subjective decision weights. The overall evaluation of a lottery L is given by:

$$V(L) = \sum_{i=1}^{k} u(x_i)\pi_i^- + \sum_{j=k+1}^{n} u(x_j)\pi_j^+ \tag{1}$$

where π^+ and π^- are the decision weights for gains and losses, respectively. The decision weights are defined as:

$$\begin{aligned}
\pi_1^- &= w^-(p_1) \\
\pi_n^+ &= w^+(p_n) \\
\pi_i^- &= w^-(p_1 + \ldots + p_i) - w^-(p_1 + \ldots + p_{i-1}) \text{ for } 1 < i \leq k \\
\pi_j^+ &= w^+(p_j + \ldots + p_n) - w^+(p_{j+1} + \ldots + p_n) \text{ for } k < j < n
\end{aligned}$$

We assume a CRRA utility function over monetary outcomes of the following form:

$$u(x) = \begin{cases} \frac{x^r}{r}, & \text{if } x \geq 0 \\ -\lambda\frac{(-x)^r}{r}, & \text{if } x < 0 \end{cases} \tag{2}$$

where $r \geq 0$ is a parameter governing the utility curvature, and $\lambda \geq 1$ is the parameter of loss aversion. Previous studies have shown that the power function fits well experimental data, for the level of monetary payoffs usually used in experimental studies (see for example Stott 2006; Balcombe and Fraser 2015; Baillon et al. 2020). When we consider losses, in order to avoid the scaling issues that a domain specific *power* (CRRA) function causes to the estimation of the loss aversion parameter, we follow Nilsson et al. (2011) and assume the same power coefficient for both gains and losses (see Köbberling and Wakker 2005; Wakker 2010; Harrison and Swarthout 2021). For the probability weighting function, we consider four specifications, the Prelec (1998) one-parameter function:

$$w(p) = \exp(-(-\log(p))^\gamma) \tag{3}$$

the Prelec (1998) two-parameter function:

$$w(p) = \exp(-(-\log(p))^\gamma)^\delta \tag{4}$$

the Tversky and Kahneman (1992) function:

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

and the Goldstein and Einhorn (1987) probability weighting function:

$$w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1-p)^{\gamma}} \tag{6}$$

All of the aforementioned specifications allow for an inverse-S shape of the weighting function, with overweighting of low probabilities and underweighting of moderate to high probabilities. The two-parameter families of the weighting function have the advantage of decomposing probability weighting to both its degree of curvature and its elevation. For instance, in Equation 6, $\delta > 0$ measures the elevation while $\gamma > 0$ measures the degree of curvature of the weighting function (likelihood insensitivity). As δ increases, the function becomes more elevated (exhibiting less overall risk aversion for gains and more for losses). On the other hand, the smaller $\gamma < 1$, the more curved the probability function, which means that the range of intermediate probabilities becomes flatter and therefore, exhibiting more rapidly diminishing sensitivity to probabilities close to the boundaries 0 and 1. As mentioned before, we allow the parameters of the weighting function to differ between domains, while we allow for both S and inverse-S shapes of the function.

To model stochastic choice we assume a logistic Luce (1959) choice rule such that the probability of choosing lottery A is given by:

$$P(A > B) = \frac{1}{1 + \exp(s(V(B) - V(A)))} \tag{7}$$

with $s > 0$ a choice sensitivity parameter, indicating how sensitively the predicted choice probability reacts to differences in the utility valuations of the two lotteries on a cardinal scale, and $V(.)$ the CPT value of the respective lottery.

In the case of lotteries defined purely in the gains domain we take into consideration the *Contextual Utility* Wilcox (2011) and the predicted probability becomes:

$$P(A > B) = \frac{1}{1 + \exp(s(\frac{V(B) - V(A)}{\nu}))} \tag{8}$$

where the normalising term ν is defined as the maximum utility over all prizes in this lottery pair minus the minimum utility over all prizes in this lottery pair. That is, the difference between the two lotteries is relative to the range of outcomes found in the lottery pair. We also note that in the gains domain the model is equivalent to the Quiggin (1982) Rank Dependent Utility model.

2.2 Cognitive Toolbox

Many theories of human cognition assume that people are equipped with a repertoire of heuristics and simplifying processes to solve the tasks they face in daily life. In the literature, this idea has been theoretically modelled with the aid of a cognitive toolbox, from which people might adaptively choose their respective strategies. In the field of judgement and decision making, this idea was pioneered by Payne et al. (1993) arguing that the decision makers are equipped with a set of strategies and select among them when faced with a decision, approach which was later extended in Gigerenzer (2002), modelling decision making as probabilistic draws from a toolbox of heuristic rules.

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 particular course of action, depending on the environment or the *ecology* of the domain upon which decisions are made. Independently of the mechanism behind the strategy selection, the outcome of this process can be modelled with the aid of a mixture proportion parameter β which indicates the probability of choosing each strategy in the toolbox. For instance, for a particular toolbox TB consisting of J strategies, each strategy f_j will be selected with probability β_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 (MAXIMIN) with probability β_1
- Avoid the lottery with the lowest payoff (MINIMAX) with probability β_2
- Pick the lottery with the highest most likely payoff (MOST LIKELY) with probability β_3
- Pick the lottery with the highest probability of the highest possible payoff (MOST PROBABLE) with probability $1 - \sum_{i=1}^{3} \beta_i$

This modelling specification allows for the underlying cognitive process of strategy selection to remain unspecified, given that the value of the parameter vector β 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 β_j:

$$p(A|TB) = \sum_{j=1}^{J} [\beta_j \times P(A|f_j)] \tag{9}$$

where $P(A|f_j)$ is the individual predicted probability of each strategy. Since the heuristics generate ordinal choice propensities (i.e. deterministic), we follow Balcombe and Fraser (2015), Rieskamp (2008) and He et al. (2022) and assume a *constant-error* choice rule to capture stochastic choice in the data[3]. The constant-error specification, has been widely used in the game-theoretical literature (i.e. trembling hand) and it has been populated by Harless and Camerer (1994) in the context of risky choice. 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 ε she makes a mistake[4]. 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. The next modelling choice we need to make, is how many heuristics to include in a toolbox. The previous literature has assumed an arbitrary, fixed number of heuristics, same for all the subjects. Nevertheless, Scheibehenne et al. (2013) discuss how restricting the repertoire to only a few strategies would ignore any *intra* and *inter*-individual differences through qualitatively different processes on the one hand, and how it can lead to the strategy sprawl problem, if one assumes too many strategies for a particular subject. In our analysis, we aim to identify the optimal toolbox for each subject, both in terms of size (how many strategies) and in terms of content (which strategies). The process we adopt is as follows. First, we adhere to Glöckner and Pachur (2012) and investigate the performance of 11 heuristics as potential components of a cognitive toolbox[5]. Then, we calculate all the potential combinations of heuristics. A set of n elements has $2^n - 1$ potential subsets when the null subset is not taken into consideration. This means that if one considers all the toolboxes of any size (ranging from toolboxes with only 2 heuristics to toolboxes with all 11 available) it gives in total 2036 potential toolboxes (excluding the null toolbox and the toolboxes with only one heuristic available). In order to reduce the number of models to estimate, we consider toolboxes of size up to 5, giving a total of 1012 toolbox models[6]. There are different types of heuristics focusing on one or multiple attributes. Some of the heuristics focus exclusively on the monetary payoffs, such as the Minimax, the Maximin or the Better than Average heuristic (outcome heuristics), while others focus on a combination of payoffs

[3]Since a heuristic choice rule predicts only deterministically, there is lack of clarity of how the deterministic prediction of the theory translates into a probability of observing one choice or the other. Andersen et al. (2010) discuss ways of how one can modify the Priority Heuristic to make it worth testing against any real data, but they conclude that any modification of this kind would be contrary to main purpose of the model. We therefore feel that the constant-error stochastic rule is the most natural one can assume.

[4]This is the part $P(A|f_j)$ in Equation 9.

[5]The full list of the heuristics along with a description of the choice mechanism behind each heuristic is provided in Table 3.

[6]In particular, with 11 available heuristics, there are 55 potential combinations for a toolbox of size 2, 165 for a toolbox of size 3, 330 for a toolbox of size 4, and, 462 for a toolbox of size 5. For example, with 4 available alternatives J with $j \in \{A, B, C, D\}$ one can form the following toolboxes: (1) nothing, (2) A or B or C or D, if the size of the toolbox is 1, (3) AB or AC or AD or BC or BD or CD, if size is 2, (4) BCD or ACD or ABD or ABC, if the size is 3, and; (4) ABCD, if the size is 4.

and probabilities (dual heuristics), such as the Least Likely, the Most Likely or the Probable (Brandstätter et al.; 2006). Finally, there are heuristics for multiple-attribute choice which include the Lexicographic (Gigerenzer and Goldstein 1996: Tversky 1972), the Priority (Hill et al. 2017; Todd et al. 1999) and the Tallying heuristic (Parpart et al. 2018; Czerlinski et al. 1999; Dawes 1979). The latter follow Rubinstein (1988) three-step model, where the agent applies an algorithmic process of decision making, going through various degrees of reason, and if two options are similar in terms of one reason (e.g. dominance) attention is shifted to other reasons (e.g. similarity). *A priori*, we expect that decision makers who are using a toolbox to decide, will delegate a small number of heuristics due to cognitive, time or other limitations. Mohnert et al. (2019) provide support in favour of this modelling choice as they find that the majority of the estimated toolboxes are of size 4, while research on model crowds indicates that a number close to five models is optimal (see Makridakis and Winkler 1983; Ashton and Ashton 1985; He et al. 2022.) and adding further models diminishes the prediction capacity of the select crowd. In addition, we do not take into consideration any toolboxes of size one, given that the aim of this paper is to relax the assumption that subjects are using a single heuristic and to avoid this kind of one-to-one comparison that previous research has investigated. In the next section we describe both the estimation and the model selection method that we adopt.

3 Bayesian Hierarchical Modelling

There are various ways one can adopt to estimate structural decision making models. The most common approach is the use of subject level Maximum Likelihood Estimation techniques (MLE). Nevertheless, MLE may generate noisy and unreliable estimates and therefore can produce extreme estimates for some of the subjects if there is a lack of a large number of observations. In addition, MLE is susceptible to overfitting and may adjust mostly noise rather than the actual preferences of the subject, leading to very poor predictive performance of the models. An alternative way is to pool all the data together and estimate a model for the representative agent, assuming a particular preference functional. Nevertheless, several studies have provided evidence against the assumption of a single data generating process and proposed the use of finite mixture models instead (see Harrison and Rustrom 2008; Fehr-Duda et al. 2010; Conte et al. 2011; Alam et al. 2022). While a useful approach that allows to test the presence of more than one preference functionals, they rely on the extreme assumption of the presence of n representative agents, one for each assumed preference functional (that is for instance, all EUT subjects share the same behavioural parameters, all CPT the same and so on).

Scheibehenne et al. (2013) provide arguments of how the Bayesian formalism can allow toolbox approaches to be rigorously tested via the use of the *Bayes Factor* (Kass and Raftery; 1995), a unifying comparison metric that quantifies the extend to which data support one model over another, taking model complexity into account. To mitigate the drawbacks of MLE, we adopt hierarchical Bayesian estimation techniques (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 hierarchical Bayesian models for choice models under risk and Stahl (2014) for ambiguity models). The key aspect of hierarchical modelling is that even though it recognises individual variation, it also assumes that there is a distribution governing this variation (individual parameter estimates originate from a group-level distribution). A hierarchical Bayesian model simultaneously estimates the individual level parameters, along with the hyper-parameters of the group level distributions. In typical hierarchical models, the estimates of the low level parameters are pulled closer together than they would in the absence of a higher-level distribution, leading to the so called *shrinkage* of the estimates.

As Baillon et al. (2020) highlight, Bayesian Hierarchical modelling is a compromise between a representative agent and subject-level type estimation. It estimates the model parameters for each subject separately, but it assumes that subjects share similarities and draw their individual parameters from a common, population level distribution. In that way, individual parameter estimates inform each other and lead to a *shrinkage* towards the group mean that reduces biases in parameter estimates.

We follow Rouder and Lu (2005) and Nilsson et al. (2011) set-up and we estimate all the specifications using BHM. Each subject i made a series of N binary choices in a given dataset and the observed choices vector is denoted by $D_i = (D_{i1} \cdots D_{iN})$. Every subject is characterised by its own parameter vector B_i and we assume that all the parameters are normally distributed $(b_i \sim N(\mu_b, \sigma_b))$, while for the hyper-parameters we assume normal priors for the mean μ_b and uninformative priors (uniform) for σ_b. For the mixture parameter vector β in the toolbox model, since it represents a probability distribution and the parameters in β are not independent from each other, we assume that it is a J-dimension categorical variable which follows a Dirichlet distribution $\beta \sim Dirichlet(\pi)$ with π a diffuse hyper prior parameter for the distribution. We also follow the standard procedure and transform all the parameters to their exponential form to ensure that they lie within the appropriate bounds.

The likelihood of subject's i choices is given by:

$$P(D_i|B_i) = \prod_{n=1}^{N} P(D_{i,n}|B_i)$$

where $P(D_{i,n}|B_i)$ is the predicted probability for each lottery pair n, as this was presented in the previous section. Combining the likelihood of the observed choices and the probability distribution of all the behavioural parameters, the posterior distribution of the parameters is given by:

$$P(B|D) \propto P(D|B) \times P(B)$$

with $P(D|B)$ being the likelihood of observed choices over all the subjects and $P(B)$ the priors for all parameters in the set B. Monte Carlo Markov Chains (MCMC) were used to estimate all the specifications. The estimation was implemented in JAGS (Plummer; 2017). The posterior distribution of the parameters is based on draws from two independent chains, with 50,000 MCMC draws each. Due to the high level of non-linearity of the models, there was a burn-in period of 25,000 draws, while to reduce autocorrelation on the parameters, the samples were thinned by 10 (every tenth draw was recorded). Convergence of the chains was confirmed by computing the $\hat{R}$ statistic (Gelman and Rubin; 1992).

All the inference and the subsequent comparison of the models is based on the log *Bayes Factor* measure (Kass and Raftery; 1995). Bayes factors are penalising models with a large number of parameters, prevent over-fitting, and are a good measure of the forecasting capacity of each model. The Bayes Factor is defined as $\exp(LML_{FULL} - LML_{EUT})$, where LML_i denotes the log-marginal likelihood of model i. To estimate the log-marginal likelihoods we use the Newton-Raftery estimator (harmonic mean of the log-likelihood over all draws after the burn-in period, Newton and Raftery 1994).

4 Data Sets

We estimated the models using a wide range of data sets from experimental studies covering all domains of gains, losses and mixed gambles, as well as cognitive loaded environments such as time pressed decision making and complexity. Table 1 provides a summary of all the data sets we use for our analysis. All data sets were chosen in a way that they would satisfy the following criteria:

- The experimental designs have been developed with the objective to estimate structural econometric models and involve a substantial number of tasks per participant.
- All studies have been incentivised in monetary terms.
- The studies incorporate a wide range of probabilities and outcomes, introducing various levels of complexity.

Data Set	Abbreviation	Domain	Cognitive Load	No. of Subjects	No. of Tasks	Payoff Structure	Probability Structure	Complexity Levels
Baillon et al. (2020)	BBS20	G	N	139	70	1-4 outcome Random	Random	Medium
Hey and Orme (1994)	HO94	G	N	80	100	3 outcome Fixed	Constant Multiples of 0.125	Low
Glöckner and Pachur (2012)	GP12	G/M/L	N	66	138	2 outcome Random	Random	Medium
Harrison and Swarthout (2021)	HS21	G/M/L	N	175	100	1-3 outcome Fixed	Based on the Marschak-Machina triangle	Low
Olschewski and Rieskamp (2021)	$OR21_{TP}$	TP	Y	60	150	2-4 outcome Random	Random	High
Moffatt et al. (2015)	MSZ15	Complexity	Y	80	54	3-27 outcome Fixed	Combination of simple lotteries with $p \in \{0.2, 0.3, 0.5\}$	High

Table 1: Summary of the Data Sets. G indicates the gains, L the losses and M the mixed domain. TP stands for the time pressure condition.

In the gain domain, we analyse experimental data from two studies; Hey and Orme (1994) and Baillon et al. (2020). Hey and Orme (1994) involves 80 subjects deciding over 100 pairwise choice questions. There are 3 outcomes in each of the lotteries, and these outcomes are held fixed at either £0, £10, £20 and £30, whilst the associated probabilities are all multiples of 0.125. Baillon et al. (2020) data include the choices of 139 participants from 70 binary lotteries. Each option in a lottery (Option A or B) has between one and four possible outcomes (all framed as gains). The payoffs and probabilities in their experiment have been carefully chosen to maximise statistical efficiency and minimise redundancy. On top of allowing for a wide range of number of outcomes and magnitude of probabilities, the authors ensured that within each choice pair there were non-matching maximal or minimal outcomes, questions had similar expected value, and finally, questions were orthogonal to maximise statistical efficiency. It can be argued that the tasks in Hey and Orme (1994) are relatively less complex than those in Baillon et al. (2020), as the fixed probabilities and payoffs permit for a simpler design and thus a less computationally challenging cognitive process. Furthermore, the way the maximum and minimum outcomes have been chosen in Baillon et al. (2020)[7] that may generate the appropriate *ecology* for an adaptive toolbox.

In the loss domain, we use the data from Glöckner and Pachur (2012) and Harrison and Swarthout (2021) for our analysis. Glöckner and Pachur (2012) include the choices of 66 subjects in 138 pairwise choice problems, where 70 are gains, 30 are losses, and the rest are mixed. The problems involve binary two-outcome lotteries. The tasks were a combination of lottery pairs that have been used in previous studies to capture various decision making phenomena with tasks being either randomly generated, designed to differentiate between the priority heuristic and CPT, designed to measure risk attitudes using the Holt and Laury (2002) task, or designed to measure loss aversion. They generated two sets of 138 tasks and participants faced either of each sets in two separate sessions. Here we are using the data from session 1. Harrison and Swarthout (2021) include data from 175 undergraduate students in 100 binary lotteries framed as gains, losses, and mixed outcomes. 16 of the tasks are mixed and 16 are losses, the rest are gains. Each task consists of lottery pairs with either 1, 2 or 3 outcomes, where the outcomes are fixed and the probabilities have been chosen using the indifference curves in the Marschak-Machina triangle, ensuring maximal discrimination between Expected Utility Theory and CPT.

We then include in our analysis data from environments that involve increased levels of cognitive load. For our first cognitive loaded domain, that of complexity, we analyse data

[7] See Appendix B in Baillon et al. (2020) for a description of how subsets of the task set differ in terms of the pairs' maximum and minimum outcomes.

from Moffatt et al. (2015). Their experiment involves 80 subjects making pairwise choices in 54 binary lottery tasks designed to separate between complexity aversion and risk aversion. The lotteries vary in complexity, where some involve as little as 3 outcomes, whilst the maximum includes 27 outcomes, labelling them as simple, complex and very complex. Adopting an appropriate procedure, they transformed 3-outcome simple lotteries to complex ones, and from these complex lotteries they could generate very complex lotteries, based on a similar procedure. They generate in total 27 lottery pairs (combinations of simple, complex and very complex lotteries) and they present them to their subjects twice, to test consistency. Here we use the first 27 tasks. All lotteries have the same expected value. For our second cognitive loaded domain, being decision making under time pressure, we make use of Experiment 2 from Olschewski and Rieskamp (2021). 60 subjects faced a battery of binary lotteries in two treatments, no time pressure (NTP) and time pressure (TP). We analyse data from 150 tasks, where 75 of these tasks fall in the NTP domain, and the other 75 were under time pressure. The set of lotteries was identical for the two treatments, but both the order of tasks and the positioning (left-right) were shown in a randomised order, different for each participant. The experimental design controls for the level of *complexity* of the lotteries by manipulating the number of outcomes of a gamble. There were three conditions in total: complex, where both gambles consisted of four outcomes each, safe-easy: where the safer gamble with lower variance had only two outcomes; and risky-easy, where the riskier gamble with higher variance had only two outcomes. Both outcomes and probabilities varied with a random structure. In the time pressure condition, the average time constraint was set to 4.12 seconds based on reaction times of participants from a practice experiment. Whereas during the NTP condition, participants had 30 seconds per lottery to make their decision

4.1 Complexity Index

In this subsection, we devise an index to explore the relationship between the degree of complexity of the tasks in a particular study and the percentage of subjects classified as Toolbox decision makers in that study. To this end, we develop a measure that will help us characterise the degree of *complexity* of a given dataset. In the literature, complexity of a task is mostly characterised 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 2016, Zilker et al. 2020, Fudenberg and Puri 2022). Nevertheless, this measure ignores the role of the set of attributes of each lottery (or pair of lotteries) that may affect the degree of complexity a decision maker perceives, and complexity can be seen as increasing with the number of alternatives and the number of attributes. Recently, researchers have started to include further attributes of a lottery as indicators of complexity. For instance, Diecidue et al. (2015) accounts for focusing on both the number of outcomes and the format of the probabilities in a task, distinguishing between simple (rounded) and complex (non-rounded) probabilities. Huck and Weizsäcker (1999) examine how several features of a decision task lead subjects to deviate from expected value maximisation, including the number of outcomes, the format of probabilities and outcomes, as well as distribution moments of the lotteries such as the mean and the variance, while in a similar context, Enke and Shubatt (2023) construct complexity indices by evaluating the effectiveness of numerous features in forecasting the error rate when identifying the lottery with the highest expected value in their experiment.

In our index we aim to take into consideration all the features of the decision tasks that may increase the decision maker's perception of complexity, along with the amount of cognitive load. On top of the number of outcomes and probabilities, we want to take into consideration the presentation format of probabilities and the degree of similarity between lotteries. Following Huck and Weizsäcker (1999) we include in the index the following features:

- Average number of outcomes across all lotteries ($avg\ \#outc$).
- Average number of outcomes with non-rounded probabilities (non-divisible by 0.05, $avg\ \#probs$).
- Average expected value difference between lottery pairs ($avg(EV_{diff})$).
- Average standard deviation difference between lottery pairs ($avg(st.dev_{diff})$).
- Ratio of unique outcomes over total number of tasks ($\#\ outcomes/\#tasks$).

We define the index I_{cplx} as:

$$I_{cplx} = avg\ \#outc + avg\ \#probs - avg(EV_{diff}) - avg(st.dev_{diff}) + \frac{\#outcomes}{\#tasks} \tag{10}$$

Note that the expected value and the standard deviation differences enter the index with a negative sign. The closer the expected value of two lotteries, the harder is to make a decision. Likewise, comparing lotteries with similar variance makes the task of identifying the riskier lottery harder. Therefore, large differences in expected value and variance reduce the overall complexity score. Finally, experimental designs that involve a large number of tasks, with non-fixed multiple outcomes are expected to be more cognitively demanding. The index captures this aspect via the ratio of the total number of unique outcome values in a dataset over the total number of tasks in this dataset. We normalise the index in the interval $[0, 1]$, with 1 indicating the highest level of complexity[8].

It is possible to assign different weights to each of the attributes, allowing some to exert a more substantial influence on the overall measure of complexity. Nevertheless, estimating these weights would necessitate data on the perceived complexity of a lottery (or a set of lotteries) based on various attributes. Gathering such data might be challenging due to its highly subjective nature[9]. As for our approach, we prefer to treat the effect of the various attributes to complexity perception as exogenous. By doing so, we employ a metric that assigns equal weight to all the attributes.

5 Results

This section presents the results of our analysis across three primary areas. Firstly, we determine whether subjects are better characterised as CPT or adaptive Toolbox decision makers. Next, among those classified as toolbox users, we ascertain the number of strategies present in their adaptive toolbox. Finally, we identify the specific heuristics that dominate the strategic portfolio of our toolbox users.

5.1 CPT Vs Heuristics

For the Hey and Orme (1994) data, we find that only 1 out of 80 subjects (1.25%) of subjects were characterised by an adaptive toolbox. The nature of the tasks involved may provide an explanation to this extreme result. Due to the way the tasks where chosen (fixed outcomes) and the nature of heuristics, there are several ties between two lotteries, where the heuristic predicts indifference, which dramatically decreases the model's predictive capacity. This may also explain the contradictory results of (Stahl; 2018)[10].

[8]While time pressure increases subjects' cognitive load, there is no seamless way to capture this feature in the index. A potential way would be to use the median response time as an indicator, expecting that low response times correspond to high time pressure. Nevertheless, given that there is no straightforward relationship between the two, we instead extrapolate the value of the index for the time pressure data. More specifically, since the tasks in Olschewski and Rieskamp (2021) are identical for both the TP and the NTP conditions, and since time pressure is expected to put high cognitive demands, we set the index for the TP condition to be equal to the mean between the NTP index and 1.

[9]Enke and Shubatt (2023) adopt a similar approach and develop an index for choice complexity using data from an experiment in which subjects are asked to identify the lottery with the highest expected value. Their index captures the predicted error rate in identifying the lottery with the highest expected value where the predictions are computed as convex combinations of choice set features. They find that the most important features include the excess dissimilarity of the lotteries, the number of outcomes, the lack of dominance, the presence of compound probabilities and the expected value difference.

[10]In fact, for 40/100; 73/100 and 35/100 of the tasks for the 1st, 2nd and 3rd tool respectively, for the HO dataset, there is a predicted probability of 0.5, as the highest, lowest or highest most likely payoffs of the two lotteries coincide. Furthermore, this dataset is known in the literature to favour Expected Utility. As Harrison and Swarthout (2021) highlight, the battery of lotteries deliberately avoided sets of lottery pairs that had generated"knife-edge"

Of the remaining 79 subjects (98.75%) who are characterised by CPT, we find that the GE weighting specification was the best performing for the majority of the CPT subjects (46%). For our the second gain-domain dataset (Baillon et al.; 2020), we find the number of participants utilising an adaptive toolbox is substantially larger than in the Hey and Orme (1994) data, with 58 out of 136 (43%) being characterised by a toolbox. Although CPT still dominates the decision-processes of these individuals, the complexity, and subsequent increased cognitive load, associated with the tasks involved may explain the increase in the toolbox percentage.

The results from the loss domain are fairly consistent across data sets, with 26% of subjects in the Harrison and Swarthout (2021), and 20% of subjects in the Glöckner and Pachur (2012) fitting to an adaptive toolbox over CPT. Interestingly, of the remaining subjects in both studies who were better characterised by CPT, 80% in the Glöckner and Pachur (2012) were best fit by a GE weighting function, and 72% in the Harrison and Swarthout (2021) were best fit by the PRL_2 weighting function. There is some imbalance regarding the number of the tasks in the gains and losses domain in the two datasets we are using. In particular, for the Harrison and Swarthout (2021) dataset, there are 16 tasks on the losses, compared to 68 on the gains domain, while in the Glöckner and Pachur (2012) data there are 30 losses lotteries against 70 in the gains domain. Therefore, it is not clear whether the difference in the performance of the two models is because of the presence of losses, or due to other reasons. To disentangle the effect of negative payoffs (losses) on behaviour, we repeat the same exercise estimating the models solely on the gain and the loss domain, and compare them within domains. For the Harrison and Swarthout (2021) data, we find that for 21.7% (30.3%) of the subjects, the cognitive toolbox best explains their behaviour in the gains (losses) domain, while for the Glöckner and Pachur (2012) data, 9.3% (10.1%) are best characterised by a toolbox model in the gain (loss) domain.

For our more cognitively demanding domains, we find a more pronounced leap towards the use of adaptive toolboxes. From the results of Olschewski and Rieskamp (2021), we see the number of individuals using a toolbox of heuristics rises from 27% in the NTP control, to 48% when they are restricted by time constraints. Therefore in the NTP condition, 44 subjects (73.3%) are characterised by a CPT specification, while in the TP condition, CPT can only successfully describe the behaviour of 31 subjects (51.7%). The time pressure in this situation has hindered the ability of individuals to subjectively weight utilities and probabilities, thus incentivising the use of heuristics to make their decisions. Finally, from the results of Moffatt et al. (2015), we see a significant switch in decision-processes, as this experimental design, consisting of extremely complex decision-tasks, meant 68% of individuals are better characterised by

tests of EUT. Their design mantra was to be agnostic about choice patterns, and see which models best characterized the data, rather than selecting lottery pairs designed to be hard for EUT per se.

Dataset	Domain/Environment	Toolbox	PRL1	PRL2	TK	GE	TOTAL
Hey and Orme (1994)	Gains	2	11	14	16	37	80
%		0.025	0.140	0.180	0.200	0.463	
Baillon et al. (2020)	Gains	60	22	39	5	10	136
%		0.441	0.162	0.287	0.037	0.074	
Harrison and Swarthout (2021)	Losses/Mixed	46	9	93	14	13	175
%		0.263	0.051	0.531	0.080	0.074	
Glöckner and Pachur (2012)	Losses/Mixed	13	2	8	0	41	64
%		0.203	0.031	0.125	0.000	0.641	
Olschewski and Rieskamp (2021)	Gains	16	15	19	9	1	60
%		0.267	0.250	0.317	0.150	0.017	
Olschewski and Rieskamp (2021)	Time Pressure	29	9	12	9	1	60
%		0.483	0.150	0.200	0.150	0.017	
Moffatt et al. (2015)	Complexity	54	1	4	19	2	80
%		0.675	0.013	0.050	0.238	0.025	

Table 2: Percentage of subjects classified as Toolbox or CPT per dataset.

an adaptive toolbox than any of the CPT specifications. Of the remaining 32% who are better fit to CPT, we find that the TK specification dominates, fitting the behaviour of 73% of these individuals.

Table 2 summarises the results for all the datasets while figure 1 illustrates the percentage of subjects classified as CPT or Toolbox decision maker in ascending order. It is apparent from the Figure that the two most cognitively demanding datasets (time pressure and complexity) have the highest frequency of subjects that resort to heuristics. This result indicates a correlation between the increased cognitive load and the incentive of subjects individuals to use simple rules of thumb to make their decision-making process easier. We explore this relationship in section 5.4.

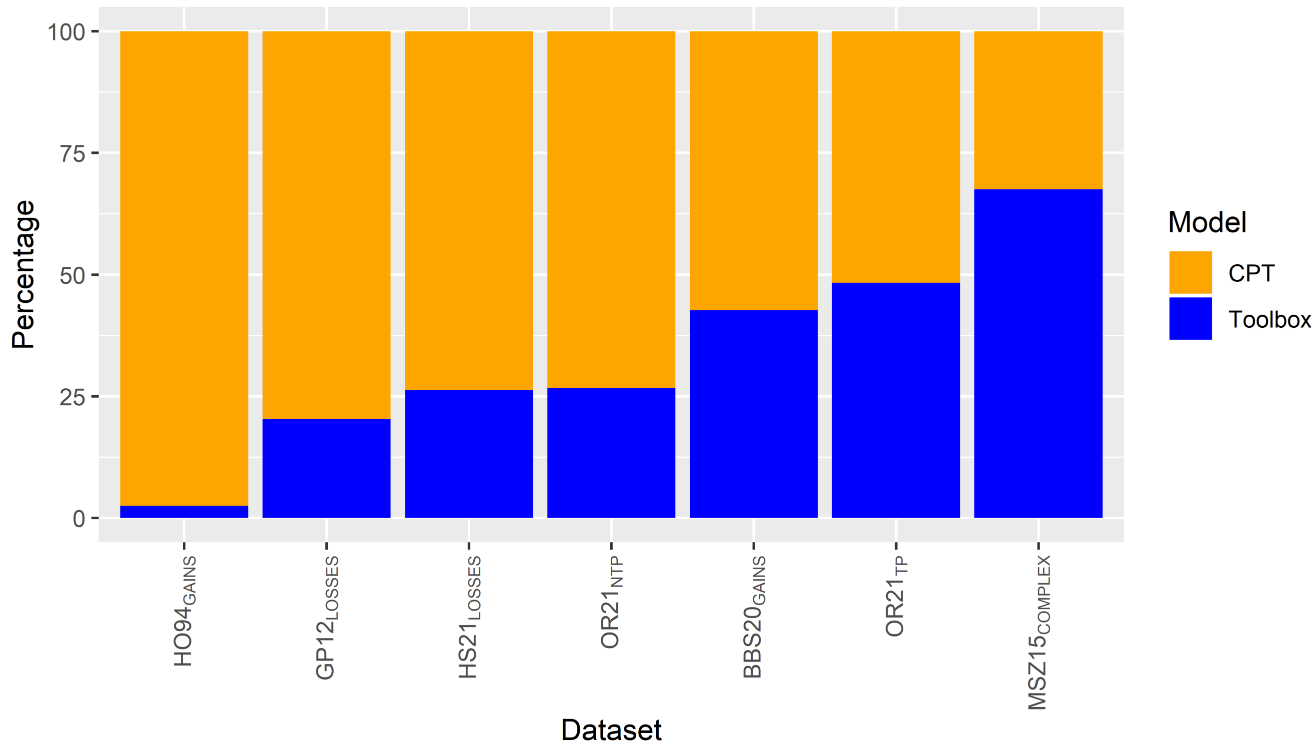


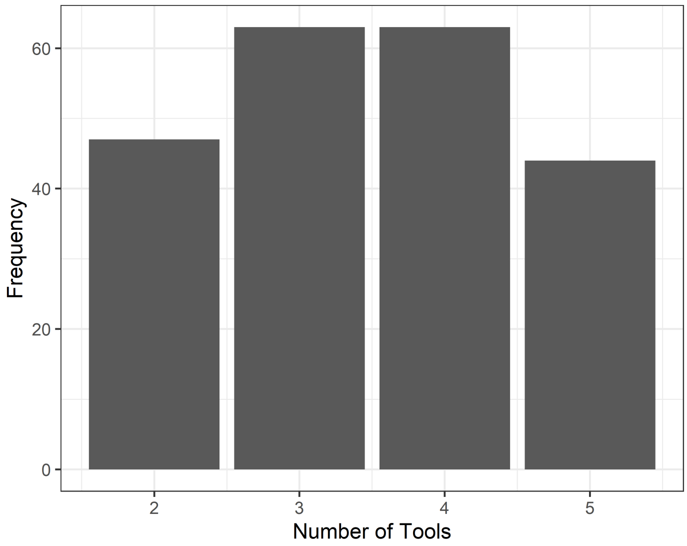
Figure 1: Percentage of subjects classified as CPT or Toolbox decision makers per dataset.

After classifying our data sets into two subdomains based on whether they are associated with an increase in cognitive load or not, we find in the low cognitive load domain that 71% of individuals are better characterised by a CPT specification, and only 29% are better described by an adaptive toolbox. Whilst in our increased cognitive load subdomain, this result changes to 50% for CPT and 50% for the toolboxes, further emphasising the increased reliance on, or switch to, toolboxes of heuristics when individuals are constrained by an increased cognitive load.

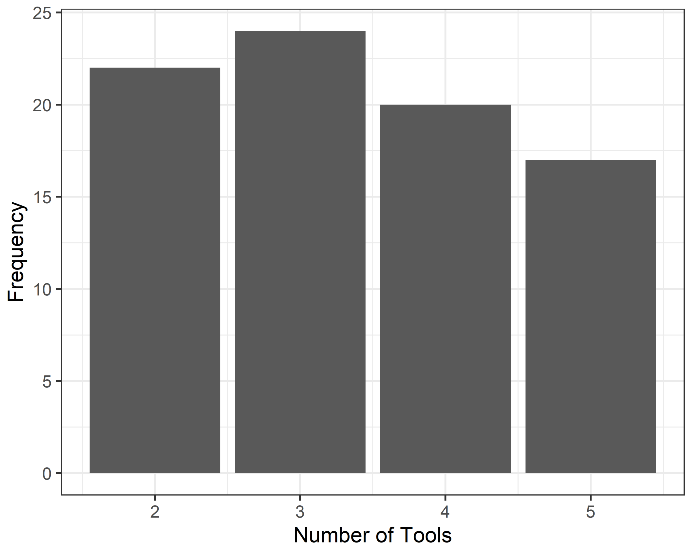
5.2 Number of strategies in a Toolbox

With regards to the optimal number of strategies to include in a toolbox, Figure 2 shows how many tools were used by each individual over all data sets.

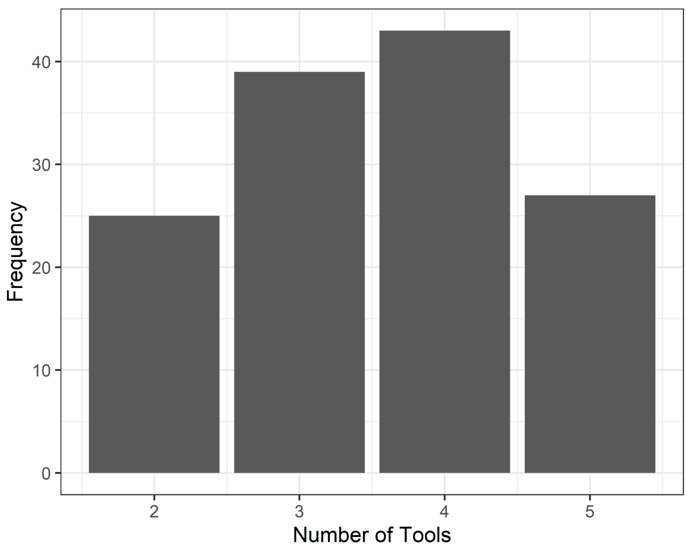
We can see that, of the 217 participants, over all data sets, who were characterised as using a cognitive toolbox, 29% had a toolbox of 4 strategies, 29% used 3 strategies, 22% used 2 strategies, and 20% used 5. We separate the analysis into our cognitive load and NCL domains to disentangle the effect that increased cognitive pressure has on the number of strategies individuals use over a series of tasks. We find that, as cognitive load increases, the number of strategies used falls. When the subjects cognitive resources are restricted, 29% of subjects use 3 strategies and 27% use 2 strategies. Only 24% and 20% of subjects use 4 and 5 strategies



(a) All



(b) High cognitive load



(c) Low cognitive load

Figure 2: Number of tools in the toolboxes. The top panel illustrates the frequency of heuristics in all toolboxes, the middle for the case of high cognitive load (time pressure and complexity) and the bottom panel for the case of low cognitive load (gains, losses and no time pressure).

respectively. On the other hand, when the tasks at hand and respective environment are not cognitively demanding, 32% of subjects use 4 strategies, 29% use 3, 20% use 5, and 19% use 2. Clearly, as the subjects cognitive capacities deteriorate, the number of strategies they can process at once declines, and they are forced to comply with a mere couple of strategies for all decision-tasks.

From a methodological perspective, our results support the existing literature in that individual toolboxes tend to hold, on average, around 3-4 strategies, and that anything over 5 would diminish the predictive capacity of the model (Mohnert et al. 2019, Makridakis and Winkler 1983, Ashton and Ashton 1985, He et al. 2022).

5.3 Which Heuristics are used

Finally, for all participants that are characterised as toolbox users, we extracted which specific heuristics formed each of their toolboxes to decipher which strategies were the most utilised in various situations. A graphical representation of our findings are represented in Figure 3. Of the 755 times that heuristics are used over decision-tasks and subjects, we find that the Minimax was used the most frequently accounting for 99 (13%) of these decisions. The Least Likely heuristic, Better-than-Average, and Equiprobable, also accounted for a substantial amount of individual decision processes. On the other hand, the Most likely heuristic, Maximin and Tallying were relied upon the least amount of times, with the latter only showing up in 6% of these toolboxes. The observed extensive heterogeneity in heuristics used, however, follows a logical pattern, with the vast majority resorting to heuristics that provide a safety net, highlighting risk aversion, and with the strategies associated with risk seeking behaviour being avoided. These results highlight the fact that adaptive toolboxes of heuristics are still able to capture complex risk preferences.

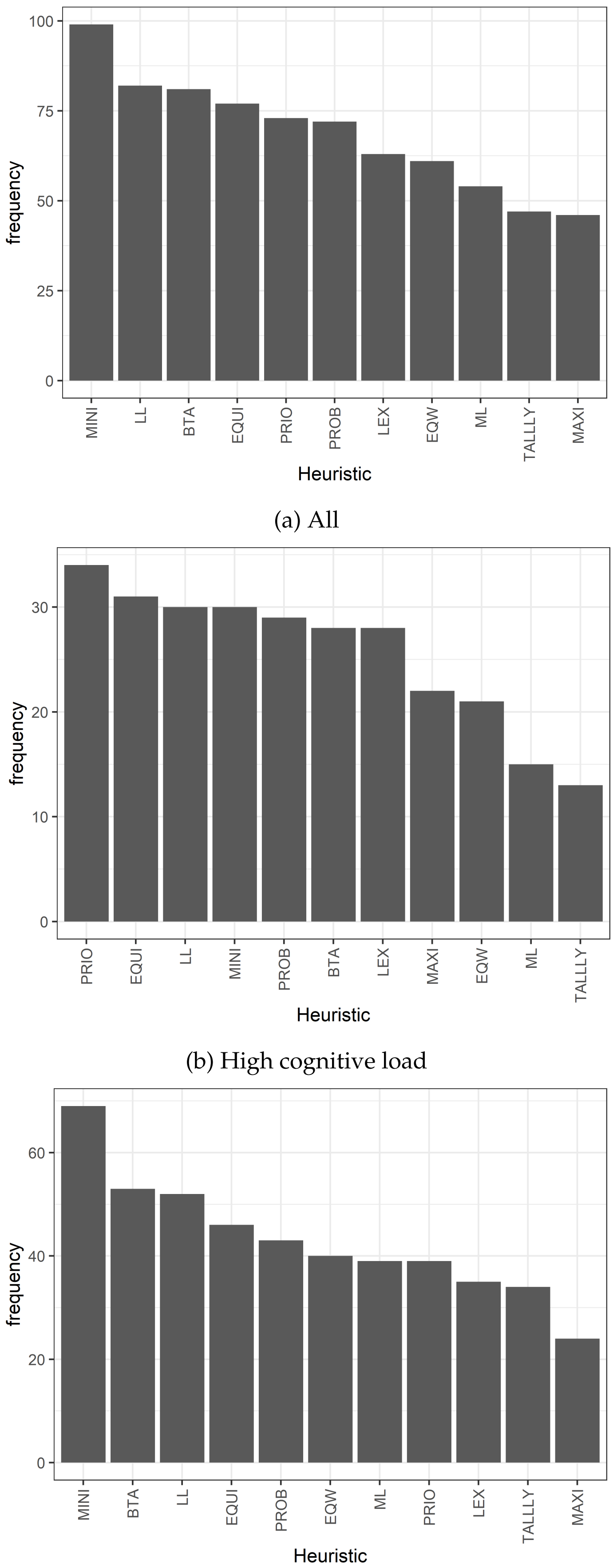
Figure 3 divides these results into two mains: those concerning an increase in cognitive load and those associated with low cognitive load. This is to depict the effect an increase on ones' cognitive load, through time pressure or increased complexity, has on the strategies adopted. There are two specific results in this domain we would like to highlight, regarding The Minimax heuristic and the Priority Heuristic. When pooling the data from the experiments involving less cognitively demanding tasks, we see that the Minimax heuristic prevailed, accounting for 69 (15%) of decisions. In this subcategory we also see that the Priority heuristic is only the 6th/11 most commonly used heuristic, accounting for only 39 (8%) of heuristic strategies in this domain. However, as time pressure is introduced and the complexity of decision-tasks increase, we find that the use of these strategies switch places with regards to their dominance in individual decision-processes. In the high cognitive load domain, we find that the use of the Priority heuristic rises to 12% (34 out of 281 cases), making it the most used strategy in this domain, whilst the use of the Minimax heuristic falls to almost 10%. Clearly the change in environment and increasing pressure on individuals cognitive resources has encouraged a change in strategy.

Finally, we separate our results into the "types" of heuristics used in the various domains, namely into three domains. The first being monetary payoff heuristics, which consist of the equiprobable, equal-weight, better-than-average, minimax, and Maximin heuristics. The second being dual heuristics, consisting of the least likely, most likely, and probable heuristics. The final domain being heuristics with a lexicographic nature, consisting of the priority heuristic, the lexicographic heuristic (multiple attribute choice), and the tallying heuristics.

Overall 48% of heuristics used were monetary payoff heuristics, 28% are dual heuristics, and 24% are the multiple attribute choice heuristics. We find that this result remains the same across the cognitive load and non-cognitive load domains. Finally, once we account for the fact that we include more monetary payoff heuristics than its counterparts and adjust our results so they are directly comparable, these results change to 36% for monetary payoff heuristics, 34% for dual heuristics, and 30% for heuristics of multiple attribute choice. This result strengthens our hypothesis that in the analysis of individual decision-making using adaptive toolboxes, it is crucial to accommodate toolbox heterogeneity and consider a diverse range of strategies. It is evident that individuals rely on various heuristics, spanning different categories, when making economic decisions.

5.4 Complexity Index

Figure 4 displays the relationship between the complexity index for each dataset and the percentage of subjects classified as Toolbox decision makers in that dataset. As expected, datasets with fixed outcomes, rounded probabilities and high differences in expected values such as Hey and Orme (1994) or Harrison and Swarthout (2021) generate the lowest complexity score, while datasets with a large number of outcomes and zero difference in the expected value of the lottery pairs, are the most complex (i.e. Moffatt et al. 2015).



(a) All

(b) High cognitive load

(c) Low cognitive load

Figure 3: Frequency of Heuristics in the toolboxes. The top panel illustrates the frequency of heuristics in all toolboxes, the middle for the case of high cognitive load (time pressure and complexity) and the bottom panel for the case of low cognitive load (gains, losses and no time pressure).

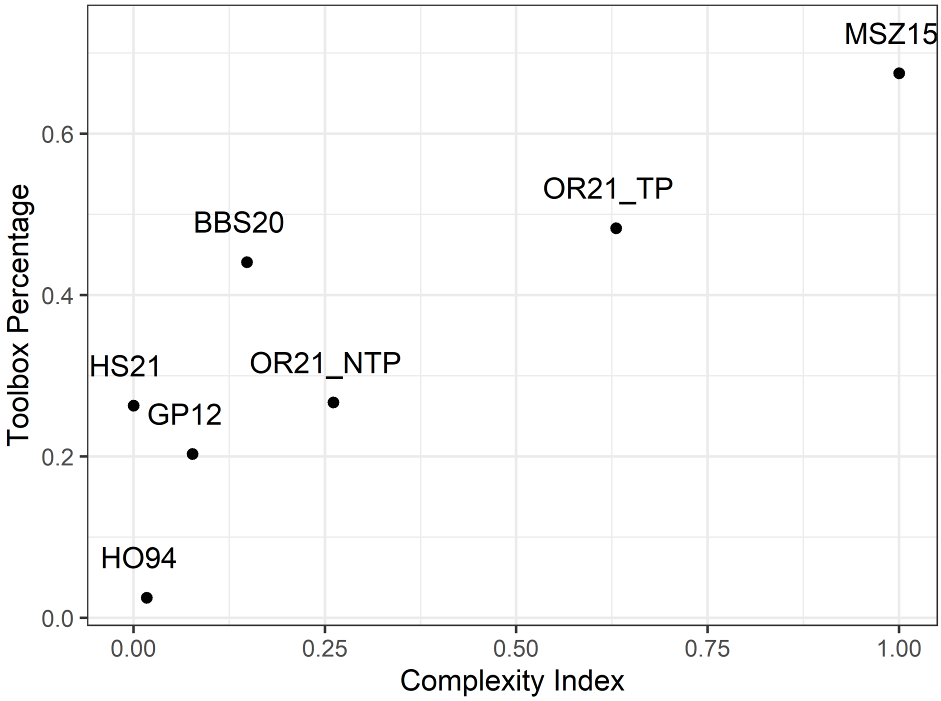


Figure 4: Complexity index per dataset

The Figure illustrates the positive relationship between complexity and use of heuristics. Using a Pearson product-moment correlation test, we find that the correlation is positive ($\rho = 0.859$) and significant (p=0.013). The latter has implications on the experimental design and stimuli used in various studies, to identify and estimate preference functionals. While there is a large literature aiming to develop methods that increase the informative content of experimental data (see for example adaptive experimental designs) there is a risk that over-complicated designs encourage participants to resort to heuristic decision making and therefore, lead to failed efforts to identify and estimate the underlying assumed preferences.

6 Conclusion

Incorporating environmental dependence into decision-making analysis is crucial for a comprehensive understanding of individual decision-making processes. Given that most day-to-day financial and economic decisions involve varying contexts and environments, it might be unreasonable to suggest that individual strategies and preferences remain constant. Specifically, when examining contextual environments demanding greater cognitive effort—whether due to increased complexity or time pressure in decision making, we propose that the nature of these decisions may overload an individual's cognitive capacity, potentially impeding the optimality and rationality of their choices. Using data from a wide range of studies encompassing diverse decision domains and environments, we investigate the efficacy of heuristics by estimating and rigorously testing toolbox models of cognition to explain behavior. Based on our complexity index, our analysis affirms that as tasks become objectively more complex, leading to an increase in choice-related cognitive load, subjects transition from the use of sophisticated expectation-based utility models to depending on a set of simplification heuristics when making decisions. Our results also indicate that with increasing task complexity, individuals use smaller toolboxes with fewer heuristics to make decisions. This is likely the result of a mental shortcut strategy to mitigate complexity, restricting the number of processable strategies an individual can handle. Lastly, we demonstrate that individuals employ a combination of various types of heuristics with extensive heterogeneity between subjects. Therefore, when assessing the capacity of toolbox models of cognition, it is crucial to incorporate all kind of heuristics, including monetary payoff heuristics, dual heuristics, and multiple attribute choice heuristics. Our analyses further support the findings of Stahl (2018): when one is agnostic on what is inside the toolbox, then she would be better off by assuming an expectation utility model, rather than a toolbox model with arbitrarily chosen heuristics. While we concentrate on just two domains that may constrain an individual's cognitive capacity, everyday decisions can be affected by numerous factors. These factors include increased time pressure, heightened uncertainty, vulnerable emotional states, interruptions or distractions, or simply a lack of experience. All of these elements can modify an individual's strategic process, ultimately leading to changes in their decisions and outcomes.

From a methodological point of view, we provide the tools on how to efficiently estimate toolbox models via Bayesian Hierarchical Modelling. This approach permits us to combine the beneficial elements of Maximum Likelihood estimation and representative agent assumptions (pooling), whilst eliminating the potential of overfitting and allowing for heterogeneity in individual preferences and strategies. It also accounts for the fact that humans share behavioural similarities, which should be exploited to enhance our understanding of generic human behaviour. As a by-product of this research, we estimate four full specifications of CPT and run a horse race comparison between different weighting functions in a wide range of decision environments. Our results indicate that the two-parameter family of probability weighting functions was always ranked best, with the PRL_2 having the best performance, followed by the GE function[11]. A final contribution is that we provide a measure of complexity of a dataset

[11] While there has been some limited research trying to identify the best combination of utility, probability weighting and stochastic link functions in the gains domain (see Stott 2006; and Balcombe and Fraser 2015 using the same dataset), there is a lack of a similar comparison in the domain of losses. Here we provide some initial evidence.

and we quantify the relationship between usage of toolboxes and complexity, finding a significant and positive relation between the two. The implications for the design of economic experiments appear to be multifaceted. Depending on the specific experimental design, it is possible that heuristics may be encouraged. It is not entirely clear-cut whether simpler designs inherently discourage heuristic-based problem-solving, nevertheless, it seems that when confronted with complexity, potential losses, or increased cognitive effort, subjects may tend to simplify their decision-making process and resort to heuristics.

Our study also adds to the burgeoning literature on collective model wisdom (model crowds) in decision analysis. As He et al. (2022) highlight, a successful application that harnesses collective model wisdom in decision analysis is Scheibehenne et al. (2013) who formulate the metaphor of the heuristic toolbox in a hierarchical Bayesian framework and show that by incorporating multiple heuristics, the toolbox explains behavioral data better than a single heuristic. Our study is the first to apply this framework to the field of risky choice.

This raises intriguing avenues for future research, including the exploration of decision environments beyond those considered here, such as decisions from experience, or choice under ambiguity, to gain a more comprehensive understanding of how the interplay between experimental design, cognitive load, and heuristics influencing decision-making.

Table 3: Table of heuristics

	Heuristic	Description
1.	Priority Heuristic	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). For loss gambles, the heuristic remains the same except that "gains" are replaced by "losses". For mixed gambles, the heuristic remains the same except that "gains" are replaced by "outcomes".
2.	Equiprobable	Calculate the arithmetic mean of all outcomes for each gamble. Choose the gamble with the highest mean.
3.	Equal-weight	Calculate the sum of all outcomes for each gamble. Choose the gamble with the highest sum.
4.	Better than average	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.	Tallying	Give a tally mark to the gamble with (a) the higher minimum gain, (b) the higher maximum gain, (c) the lower probability of the minimum gain, and (d) the higher probability of the maximum gain. For losses, replace "gain" with "loss" and "higher" with "lower" (and vice versa). Choose the gamble with the highest number of tally marks.
6.	Probable	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.
7.	Minimax	Choose the gamble with highest minimum outcome.
8.	Maximin	Choose the gamble with the highest outcome.
9.	Lexicographic	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.
10.	Least likely	Identify each gamble's worst outcome. Then choose the gamble with the lowest probability of the worst outcome.
11.	Most likely	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).

References

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

Andersen, S., Harrison, G., Lau, M. and Ruström (2010). Behavioral Econometrics for Psychologists, *Journal of Economic Psychology* **31**(4): 553–576.

Arieli, A., Ben-Ami, Y. and Rubinstein, A. (2011). Tracking decision makers under uncertainty, *American Economic Journal: Microeconomics* **3**(4): 68–76.

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.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*, Springer-Verlag New York.

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

Conte, A., Hey, J. D. and Moffatt, P. G. (2011). Mixture models of choice under risk, *Journal of Econometrics* **162**(1): 79 – 88.

Czerlinski, J., Gigerenzer, G. and Goldstein, D. (1999). *How good are simple heuristics?*, pp. 97–118.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making, *American Psychologist* **34**(7): 571–582.

del Campo, C., Pauser, S., Steiner, E. and Vetschera, R. (2016). Decision making styles and the use of heuristics in decision making, *Journal of Business Economics* **86**.

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 Shubatt, C. (2023). Quantifying lottery choice complexity. Working paper, CESifo Working Papers.

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.

Fehr-Duda, H., Bruhin, A., Epper, T. and Schubert, R. (2010). Rationality on the rise: Why relative risk aversion increases with stake size, *Journal of Risk & Uncertainty* (40): 147–180.

Fennema, H. and Wakker, P. (1997). Original and cumulative prospect theory: a discussion of empirical differences, *Journal of Behavioral Decision Making* **10**(1): 53–64.

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

Fiedler, S. and Glöckner, A. (2012). The dynamics of decision making in risky choice: An eye-tracking analysis, *Frontiers in psychology* **3**: 335.

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

Gal, D. and Rucker, D. (2018). The loss of loss aversion: Will it loom larger than its gain?, *Journal of Consumer Psychology* **28**.

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

Gelman, A. and Rubin, D. (1992). Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science* **7**(4): 457–472.

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

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic decision making, *Annual review of psychology* **62**: 451–82.

Gigerenzer, G. and Goldstein, D. (1996). Goldstein, d.g.: Reasoning the fast and frugal way: models of bounded rationality. psychological review 103(4), 650, *Psychological Review* **103**: 650–669.

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

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

Goldstein, W. and Einhorn, H. (1987). Expression Theory and the Preference Reversal Phenomena, *Psychological Review* (94): 236–254.

Harless, D. and Camerer, C. (1994). The Predictive Utility of Generalised Expected Utility Theories, *Econometrica* **62**(6): 1251–1289.

Harrison, G. and Rustrom, E. (2008). Risk Aversion in the Laboratory, *in* J. Cox and G. Harrison (eds), *Research in Experimental Economics*, Vol. 12, Emerald Group Publishing Limited, pp. 41–

196.

Harrison, G. and Rutström, E. (2008). Expected Utility Theory and Prospect Theory: one Wedding and a Decent Funeral, *Experimental Economics* **12**(2): 133.

Harrison, G. and Swarthout, J. (2019). Eye-tracking and economic theories of choice under risk, *Journal of the Economic Science Association* **5**.

Harrison, G. and Swarthout, T. (2021). Cumulative Prospect Theory in the Laboratory: A Reconsideration, *in* G. Harrison and D. Ross (eds), *Models of Risk Preferences: Descriptive and Normative Challenges*, UK: Emerald, Research in Experimental Economics.

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

Hey, J. and Orme, C. (1994). Investigating Generalizations of Expected Utility Theory Using Experimental Data, *Econometrica* **62**(6): 1291–1326.

Hilbert, M. (2011). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making, *Psychological bulletin* **138**: 211–37.

Hill, W. T., Raacke, J. D. and Park, J. A. (2017). Examining the priority heuristic in conditions of resource need levels, *Psychological Reports* **120**(5): 824–845. PMID: 28558547.

Holt, C. and Laury, S. (2002). Risk Aversion and Incentive Effects, *American Economic Review* **92**(5): 1644–1655.

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.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk, *Econometrica* **47**(2): 263–291.

Kass, R. and Raftery, A. (1995). Bayes Factors, *Journal of the American Statistical Association* **90**: 773–795.

Köbberling, V. and Wakker, P. (2005). An Index of Loss Aversion, *Journal of Economic Theory* **122**: 119–131.

Loomes, G., Moffat, P. and Sudgen, R. (2002). A Microeconometric Test of Alternative Stochastic Theories and Risky Choice, *Journal of Risk and Uncertainty* **2**: 103–130.

Lopes, L. L. and Oden, G. C. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and sp/a theory, *Journal of Mathematical Psychology* **43**(2): 286–313.

Luce, R. (1959). On the Possible Psychophysical Laws, *Psychological Review* (66): 81–95.

Makridakis, S. and Winkler, R. (1983). Averages of Forecasts: Some Empirical Results, *Manage-*

ment 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.

Mohnert, F., Pachur, T. and Lieder, F. (2019). What's in the Adaptive Toolbox and How Do People Choose From It? Rational Models of Strategy Selection in Risky Choice.

Newton, M, A. and Raftery, A. (1994). Approximate Bayesian Inference with the Weighted Likelihood Bootstrap, *Journal of the Royal Statistical Society, Series B* **56**: 3–48.

Nilsson, H., Rieskamp, J. and Wagenmakers, E.-J. (2011). Hierarchical Bayesian Parameter Estimation for Cumulative Prospect Theory, *Journal of Mathematical Psychology* **55**: 84–93.

Oberholzer, Y., Olschewski, S. and Scheibehenne, B. (2021). Complexity aversion in risky choices and valuations: Moderators and possible causes.

Olschewski, S. and Rieskamp, J. (2021). Distinguishing three effects of time pressure on risk taking: Choice consistency, risk preference and strategy selection, *Journal of Behavioral Decision Making* **34**(4): 541–554.

Pachur, T., Suter, R. S. and Hertwig, R. (2017). How the twain can meet: Prospect theory and models of heuristics in risky choice, *Cognitive Psychology* **93**: 44–73.

Parpart, P., Jones, M. and Love, B. C. (2018). Heuristics as bayesian inference under extreme priors, *Cognitive Psychology* **102**: 127–144.

Payne, J., Bettman, J. and Johnson, E. (1988). Adaptive strategy selection in decision making, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **14**: 534–552.

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

Payne, J. W. (1976). Heuristic search processes in decision making, *Advances in Consumer Research* .

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.

Payne, J. W., Bettman, J. R. and Luce, M. F. (1996). When time is money: Decision behavior under opportunity-cost time pressure, *Organizational Behavior and Human Decision Processes* **66**(2): 131–152.

Payne, J. W., Laughhunn, D. J. and Crum, R. (1980). Translation of gambles and aspiration level effects in risky choice behavior, *Management Science* **26**(10): 1039–1060.

Peterson, J., Borgin, D., Agrawal, M., Reichman, D. and Griffiths, T. (2021). Using Large-scale experiments and Machine Learning to Discover Theories of Human Decision-Making, *Science* **372**: 1209–1214.

Plummer, M. (2017). JAGS Version 4.3.0 User Manual, *Technical report*.

Prelec, D. (1998). The Probability Weighting Function, *Econometrica* **66**(3): pp. 497–527.

Quiggin, J. (1982). A Theory of Anticipated Utility, *Journal of Economic Behavior and Organization* **3**(4): 323–343.

Rieger, M. O., Wang, M. and Rayner, K. (2008). What is behind the priority heuristic? a mathematical analysis and comment on brandstätter, gigerenzer, and hertwig (2006), *Psychological Review* **115**: 274–280.

Rieskamp, J. (2008). The Probabilistic Nature of Preferential Choice, *Experimental Psychology: Learning, Memory and Cognition* (44): 1.

Rouder, J. and Lu, J. (2005). An Introduction to Bayesian Hierarchical Models with an Application in the Theory of Signal Detection, *Psychonomic Bulletin & Review* **55**: 84–93.

Rubinstein, A. (1988). Similarity and decision-making under risk (is htere a utility theory resolution to the allais paradox?), *Journal of Economic Theory* (46): 145–153.

Saqib, N. U. and Chan, E. Y. (2015). Time pressure reverses risk preferences, *Organizational Behavior and Human Decision Processes* **130**: 58–68.

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

Sokolowska, J. (2006). Risk perception and acceptance: One process or two?, *Experimental Psychology* **53**(4): 247–259. PMID: 17176656.

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

Stahl, D. (2014). Heterogeneity of Ambiguity Preferences, *The Review of Economics and Statistics* **96**(5): 609–617.

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

Stott, H. (2006). Cumulative Prospect Theory's Functional Menagerie, *Journal of Risk and Uncertainty* **32**(2): 101–130.

Thaler, R. H. (2000). From homo economicus to homo sapiens, *Journal of Economic Perspectives* **14**(1): 133–141.

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

Todd, P., Czerlinski, J., Davis, J., Gigerenzer, G., Goldstein, D., Goodie, A., Hertwig, R., Hoffrage, U., Laskey, K., Martignon, L. and Miller, G. (1999). *Simple Heuristics That Make Us Smart*.

Tom, S., Fox, C., Trepel, C. and Poldrack, R. (2007). The neural basis of loss aversion in decision-making under risk, *Science (New York, N.Y.)* **315**: 515–8.

Tversky, A. (1972). Elimination by aspects: A theory of choice, *Psychological Review* **79**(4): 281–299.

Tversky, A. and Kahneman, D. (1974). Heuristics and biases: Judgement under uncertainty, *Science* **185**: 1124–1130.

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.

Wakker, P. (2010). *Prospect Theory*, Cambridge University Press.

Wilcox, N. (2011). Stochastically more Risk Averse: A Contextual Theory of Stochastic Discrete Choice under Risk, *Journal of Econometrics* **162**(1): 89 – 104.

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.

Zindel, M. L., Zindel, T. and Quirino, M. G. (2014). Cognitive bias and their implications on the financial market, *International Journal of Engineering and Technology* **14**(3): 11–17.