Accounting for the Instability of Risk Preferences: Salience Theory versus Cumulative Prospect Theory *

Lunzheng Li†

Abstract

Salience theory is a powerful alternative to prospect theory in accounting for paradoxes of choice under risk. In risk choice settings where the majority of subjects exhibits unstable risk attitudes, we experimentally investigate the descriptive and predictive power of salience theory, and compare it with cumulative prospect theory. We find that both theories unsurprisingly outperform expected utility theory, which does not account for the instability of risk preferences and cumulative prospect theory outperforms salience theory by an insignificant margin. We attribute this small gap to the unsophisticated specification of the salience function and the substantial heterogeneity of the local thinking parameter. Salience theory captures important features of unstable risk preferences, yet further work on the functional representation of the theory is necessary to make it as applicable as cumulative prospect theory.

Keywords: salience theory, cumulative prospect theory, risk preference, lab experiment

JEL Classification: D81, D91

*I am deeply grateful to Zacharias Maniadis, John Hey, Michael Vlassopoulos, Stefan Trautmann, and seminar participants at the University of Southampton for thoughtful comments. This work has received financial support from the Economic and Social Research Council, UK.

†University of Southampton, Department of Economics, email: ll12g13@soton.ac.uk

1 Introduction

Expected utility theory (henceforth denoted as EUT), the standard tool with which economists used to model risk, assumes that individuals respond to risk in a consistent manner. The theory may provide a valuable normative guide, but it faces difficulties while playing the role of a descriptive theory. Mounting empirical evidence suggests that people systematically switch between risk aversion and risk-seeking, depending on the situation. Allais (1953) shows that when people choose between two lotteries, adding a common consequence to both lotteries might change the preference order, which contradicts the independence axiom. Lichtenstein and Slovic (1971) find that subjects tend to choose relative safer lotteries when making choices, but are willing to pay more for risky ones. Moreover, Kahneman and Tversky (1979, 1981) suggest that framing lotteries differently or replacing gains with losses could lead to a reversal of preference orders and other paradoxes.

In the past several decades, alternatives of EUT have been developed to explain this instability of risk preference. In this paper, we are particularly interested in Bordalo et al. (2012)'s salience theory (henceforth denoted as ST). The idea of the theory is that a decision maker's attention is drawn to salient consequences, and the probabilities are distorted accordingly. Bordalo et al. (2012) introduced the theory as a unified explanation for several anomalies related to unstable risk preferences, such as excessive risk-seeking behaviour and the Allais paradox. Importantly, the new theory can explain these phenomena using only a small set of assumptions about the function that guides how salient lottery outcomes are perceived. ST also makes new predictions which contradict prospect theory (Kahneman and Tversky, 1979). In addtion, the authors presented the results of a series of online experiments as empirical evidence.

This novel theory has attracted considerable attention from empirical reseachers. Kontek (2016) points out that the certainty equivalents of lotteries are undefined according the theory for some range of probabilities, and the theory also violates monotonicity. Nielsen et al. (2018) report an online experiment with 473 participants. They manipulate the salience value of the good and bad consequences in each lottery and the results are consistent with the prediction of ST. Frydman and Mormann (2018) replicate the Allais paradox experiment with one adjustment: they set two lotteries to be correlated. ST relies on the joint distribution of the lotteries, while other theories, such as Tversky and Kahneman (1992)'s cumulative prospect theory (henceforth denoted as CPT) do not, so Frydman and Mormann (2018) conclude that ST provides a coherent framework to understand the Allais paradox. Dertwinkel-Kalt and Köster (2019) link risk preference to the skewness of the probability distribution of lotteries, and experimentally show that ST accommodates such preferences better than CPT. Königsheim et al. (2019) focus on the local thinking parameter, which measures how much individuals' decision weights are distorted because of limited attention. They calibrate the parameter and argue that the estimate depends on whether the lottery is downside or upside salient.[1]

We report a laboratory experiment conducted to examine the empirical validity of ST. The experiments mentioned above test ST on the basis of its axiomatic fundamental and behavioural assumptions, i.e., the researchers examine whether salience affects risk taking or not and how. However, our experiment is a different exercise. The purpose of this paper is not to verify or challenge the behavioural tenets of ST, but to compare the theory to other candidate theories in the sense of empirical fitness and predictions. We apply the theory in a straightforward manner to a risk choice setting where the majority of subjects exhibit unstable risk attitudes and check how accurately the theory describes and predicts decisions. We choose the popular CPT as the baseline model for evaluation since this theory has been tested thoroughly and it is useful in terms of application (Gonzalez and Wu, 1999; Wu and Markle, 2008; Hey et al., 2010; Kothiyal et al., 2014; Georgalos, 2019). The classical EUT is also included in the final comparison. The results show that in terms of both descriptive and predictive power, CPT outperforms ST and ST outperforms EUT. It is to be expected that EUT be dominated since it does not account for unstable risk preferences. On the other hand, CPT and ST are racing closely in terms of predictive power. In future research, the gap may be reduced by improving two aspects of ST: (i) The functional representation of salience function; (ii) Deriving different local thinking parameters for lotteries with opposite salience directions.

The structure of the paper is as follows. In Section 2, we review the investigated models and introduce the functional forms used in the analysis. Section 3 introduces the experimental design. The results are presented in Section 4. Section 5 concludes.

[1]We examined the local thinking parameter in the same manner. See Section 4.4 for details.

2 Review of theories

The theories under investigation are EUT, CPT and ST. In this section, we review the theories and present the preference functional used in our analysis. We also explain how the theories can (or cannot) account for phenomena related to unstable risk preferences. Kahneman and Tversky (1979)'s version of common consequence effect example (Equation 1) is used for illustration. In their experiment, subjects are asked to choose between $L(c)$ and $R(c)$ for different values of c. It is clear that $R(c)$ is the safer option for any c. The results show that the majority of subjects choose the safer option $R(c)$ when $c = 2400$. However, most of them switch to the riskier option $L(c)$ when $c = 0$.

$$L(c) = \begin{cases} 2500, & \text{with prob.} \quad 0.33 \\ 0, & \quad\quad\quad\quad\quad\; 0.01 \\ c, & \quad\quad\quad\quad\quad\; 0.66 \end{cases} \quad ; \quad R(c) = \begin{cases} 2400, & \text{with prob.} \quad 0.34 \\ c, & \quad\quad\quad\quad\quad\; 0.66 \end{cases} \tag{1}$$

Expected Utility Theory

An EUT agent's preference order over $L(c)$ and $R(c)$ does not depend on the value of c because of the independence axiom, and her risk preference can be characterised simply by the curvature of the utility function. Our chosen utility function is:

$$U(x) = x^{\tau} \tag{2}$$

where $\tau > 0$. Note that $0 < \tau < 1$ indicates risk-averse, $\tau = 1$ indicates risk neutral, and $\tau > 1$ indicates risk-seeking preferences.

Cumulative Prospect Theory

CPT is an advanced version of prospect theory. In the basic expected utility framework, a subject is assumed to have an underlying value function and the utility is linear in probabilities. Prospect theory maintains the idea of the fixed value function, but the value is relative to a reference point. However, instead of using raw probabilities as decision weights, the theory converts the probabilities into decision weights according to a non-linear weighting function.

Kahneman and Tversky (1979) derive the weighting function based on psychological insights and the function overweights small probabilities and underweights moderate and high probabilities. In terms of the Kahneman and Tversky (1979)'s common consequence effect example, when c switches from 0 to 2400, the probability associated with the zero payoff in $L(c)$ drops significantly (from 0.67 to 0.01). Thus the probability is being overweighted, and subjects choose the safer option $R(c)$.

The shape of the non-linear weighting function plays a vital role in explaining the common consequence effect and other anomalies. However, a mere monotonic transformation of outcome probabilities causes violations of stochastic dominance and the problem of non-additivity.[2] To solve these problems, Tversky and Kahneman (1992) modified the weighting strategy by incorporating rank-dependent utility (Quiggin, 1982): Instead of transforming each probability separately, the new model transforms the entire cumulative distribution function. Considering a risky prospect which is represented as n pairs of (x_i, p_i), where x_i is the payoff and p_i is the corresponding probability, and $x_1 < x_2 < ... < x_n$, the decision weight π_i equals $w(p_i + ... + p_n) - w(p_{i+1} + ... + p_n)$. $w(.)$ is the cumulative probability weighting function, and $w(p_i + ... + p_n)$ measures the probability of getting a value at least as good as x_i. On the other hand, the features of the value function are maintained. The value function is defined on deviations from a reference point, is concave for gains and convex for losses and it is steeper for losses than for gains. With these weighting and value functions, CPT allows for unstable risk preferences, in particular, subjects become risk-averse for gains and risk-seeking for losses in high probability events, while they switch to risk-seeking for gains and risk-averse for losses in low probability events.

We follow the original parametric representation of the cumulative probability function and value function of Tversky and Kahneman (1992). Our experiment does not deal with

[2]Consider the following example. Lottery A has two outcomes: 1 and 100, and the corresponding probabilities are 0.01 and 0.99. Lottery B's outcomes are all the integers from 1 to 100, and each outcome corresponds to a probability of 0.01. Obviously, lottery A first-order stochastically dominates lottery B. However, violations of stochastic dominance may occur if the outcome probabilities are distorted in the manner of a monotonic weighting function (low probabilities get larger while high probabilities get smaller). Moreover, for lottery B, the weights obviously do not add to unity after the distortion.

losses,[3] thus the weighting function $w(p)$ is:

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}} \tag{3}$$

where $0 < \gamma < 1$. The value function is:

$$v(x) = x^\alpha \tag{4}$$

where $0 < \alpha < 1$, and we assume that the reference point for evaluation is zero.

Salience Theory

Unlike CPT or several other generalisations of EUT which focus on the shape of the probability weighting function, the key point of ST is that the distortion on probabilities depends on the payoffs of both lotteries, specifically, on "how salient a state is". Bordalo et al. (2012) refer to decision makers as local thinkers, i.e., they can only think "locally" due to limited attention or cognitive limitations. Therefore, a local thinker can only focus on (or process) the most salient state, and tends to overweight it. We present the model with a two-lottery choice set $\{L_1, L_2\}$ where both lotteries have two possible outcomes, i.e., $L_i = (x_i^1, p_i; x_i^2, 1-p_i)$ where x_i^1, x_i^2 are the possible outcomes, $p_i, 1-p_i$ are the corresponding probabilities, and $i \in \{1, 2\}$. This choice problem can be described as a set of states of the world $S = \{s_1, s_2, s_3, s_4\}$, and the four possible states $s \in S$ are presented in Table 1. State s has a payoff combination of (x_i^s, x_j^s) where $j \in \{1, 2\}$ and $j \neq i$, and a probability of π_s.

Table 1: Possible states of the world in the example

States	Payoff combination	Probability
s_1	$(x_1^{s_1}, x_2^{s_1}) \equiv (x_1^1, x_2^1)$	$\pi_{s_1} = p_1 p_2$
s_2	$(x_1^{s_2}, x_2^{s_2}) \equiv (x_1^1, x_2^2)$	$\pi_{s_2} = p_1(1-p_2)$
s_3	$(x_1^{s_3}, x_2^{s_3}) \equiv (x_1^2, x_2^1)$	$\pi_{s_3} = (1-p_1)p_2$
s_4	$(x_1^{s_4}, x_2^{s_4}) \equiv (x_1^2, x_2^2)$	$\pi_{s_4} = (1-p_1)(1-p_2)$

A salience function $\sigma(.)$ is defined over the payoff combinations and is used to measure the

[3] We choose not to deal with losses for two reasons: Firstly, ST and CPT have the same degrees of freedom if negative payments are not included. This makes the comparison simpler and more meaningful. Secondly, including losses increases the decisions subjects need to make and inevitably increases the experiment time, which was a constraint in our case.

perceived difference between the two payoffs in each state. It satisfies two conditions: ordering (two payoffs define an interval for each state, and a state is less salient if the corresponding interval is a subset of the alternative) and diminishing sensitivity (keeping the payoff difference constant, a state is less salient when payoffs lies futher from zero).[4] Bordalo et al. (2012) suggest the following continuous and bounded function for state s:

$$\sigma(x_i^s, x_j^s) = \frac{\left|x_i^s - x_j^s\right|}{\left|x_i^s\right| + \left|x_j^s\right| + \beta} \quad , (\beta > 0). \tag{5}$$

The local thinker ranks the states according to the value of the salience function, with lower k_s (k_s is a positive integer and it represents the ranking of state s) indicating higher salience, and the distorted decision weight is given by:

$$\pi_s^d = \pi_s \times \frac{\delta^{k_s}}{\sum_{r \in S} \delta^{k_r} \pi_r} \tag{6}$$

where $0 < \delta \leq 1$, and δ is the local thinking parameter which measures how much a local thinker's attention is drawn by the salient states. $\sum_r \delta^{k_r} \pi_r$ is used to normalize $\sum \pi_s^d$ to 1. Therefore, states ranked higher (the most salient states) are overweighted and states ranked lower (the least salient states) are under-weighted. Bordalo et al. (2012) assume a linear value function $v(x) = x$, and the local thinker evaluates L_i as:

$$V(L_i) = \sum_{s \in S} \pi_s^d x_i^s \tag{7}$$

Consider Kahneman and Tversky (1979)'s common consequence effect example. When the common consequence c is 0, the most salient state is associated with the payoff combination (2500, 0), which makes the riskier $L(c)$ more attractive. However, when the common consequence c becomes 2400, the most salient state becomes (0, 2400) where the payoff of $R(c)$ clearly dominates. Thus, subjects choose the safer option $R(c)$.

Several auxiliary assumptions on the basic ST framework are necessary in order to obtain reasonable parameter estimates. The ranking k_s in Equation 6 creates a discontinuity in the utility function, which gives us difficulty in estimating the salience function and it does not

[4] Bordalo et al. (2012) also mentioned a "reflection condition" to extend the theory to losses, which is not within the scope of our study.

contain the information of the magnitude of salience in distortions. Since a more salient state is associated with a smaller k_s, we replace k_s with $\frac{1}{\sigma(x_i^s, x_j^s)}$ to smooth out the utility function.[5] Also, to avoid the denominator becoming zero, we modify the salience function to the following:

$$\sigma(x_i^s, x_j^s) = \frac{\left|x_i^s - x_j^s\right| + \theta}{\left|x_i^s\right| + \left|x_j^s\right| + \beta} \quad , (\theta, \beta > 0), \tag{8}$$

without loss of generality, we set $\beta = \lambda\theta$ with $\lambda > 0$. This salience function satisfies the ordering condition indicating that $\sigma(x_i^s, x_j^s)$ increases with the gap between x_i^s and x_j^s. Therefore, for any fixed x_j^s, the function is increasing in x_i^s. In a special case when $x_j^s = 0$ and $x_i^s > 0$, we have:

$$\sigma(x_i^s, 0) = \frac{x_i^s + \theta}{x_i^s + \lambda\theta} = 1 + \frac{(1-\lambda)\theta}{x_i^s + \lambda\theta} \quad , (\lambda, \theta > 0), \tag{9}$$

to satisfy the ordering condition, λ should be larger than 1. We set λ to e (Euler's Number) for calibration purposes.[6] We use the following salience representation in the analysis:

$$\sigma(x_i^s, x_j^s) = \frac{\left|x_i^s - x_j^s\right| + \theta}{\left|x_i^s\right| + \left|x_j^s\right| + e\theta} \quad , (\theta > 0), \tag{10}$$

In general, this setting satisfies the conditions of the theory and gives us reasonable estimates of θ and δ. Regarding the value function, we stick to the linear value so that CPT and ST have the same number of parameters.

3 The Experiment

We use a series of binary choice questions to elicit risk preferences. In the experimental practice of risk preferences elicitation, researchers usually ask subjects three forms of questions: binary choice questions, reservation price questions and allocation questions (Hey and Pace, 2014).[7] Allocation questions are not our choice because of the context-dependence nature of

[5]Following Bordalo et al. (2012)'s suggestion, we replaced k_s with $-\sigma(x_i^s, x_j^s)$ at first. However, this approach is inappropriate for our case since it yields extreme estimates (both δ and β are extremely small).

[6]If λ is a rational number, for any payoff combinations which satisfy $\frac{\left|x_i^s - x_j^s\right|}{\left|x_i^s\right| + \left|x_j^s\right|} = \frac{1}{\lambda}$, $\sigma(x_i^s, x_j^s)$ is a constant which clearly violates diminishing sensitivity. The irrational number e is chosen only for its aesthetic feature and it is mathematically irrelevant.

[7]A typical allocation question design is like the following: subjects are given a fixed amount of tokens and they are asked to allocate the tokens to events with different probabilities. Subjects maximise their preference functions to make the allocation decisions.

ST. In particular, an ST agent judges a lottery differently when the alternative is different, hence the theory does not have a unified preference functional form for a single lottery, which makes the allocation question method infeasible. Accordingly, the other two forms of questions have special advantages. According to Hey et al. (2010), binary choice questions "are easier to explain to subjects; easier for them to understand; and less prone to problems of understanding associated with the various mechanisms for eliciting", while obtaining the reservation prices (certainty equivalents) of lotteries enhances the information in data. We believe that the attractions of both methods are important to our experiment. Therefore, in the first stage, we let subjects make choices between a lottery and a set of consecutive sure payoffs, so that the certainty equivalents of the lottery can be estimated if necessary.[8] In the second stage, a subject is given a series of binary choice questions and she is paid according to her decision of one randomly chosen question. The observations derived in the two stages are used for different purposes, which we shall explain later.

As mentioned above, we shall judge a theory based on its descriptive and predictive power rather than its behavioural hypothesis. We follow Hey et al. (2010) in determining which theory is 'better'. Briefly speaking, subjects are required to answer two sets of binary choice questions, with part of the observations used for calibration, and the remaining part used for predictive capacity testing. In particular, predictive capacity can be measured by predicted log-likelihoods, and the theory which has larger predicted log-likelihoods is considered to be outperforming the others. We design our experiment in two stages: Stage 1 is used for calibration and testing descriptive power, and Stage 2 is used for testing predictive capacity. Stage 1 questions are designed to give enough information for calibration. Stage 2 questions focus on two typical phenomena which exhibit unstable risk attitudes: risk-seeking behaviour (typically a risk-averter might prefer a relatively small chance of winning a big prize to the expected value of such lottery) and the common consequence effect (the preference order over two lotteries is affected by the change in a common consequence).

[8] Subjects' certainty equivalents or reservation prices are not elicited directly. However, this design gives us enough information to infer them.

Stage 1

In Stage 1, each subject faces 36 rounds, and in each round, she needs to make decisions between a risky prospect and a sure outcome eight times. This design allows us to get as much information as we would by asking subjects for their reservation prices. In a typical round, the screen displays a lottery on the left and a descending series of eight sure outcomes on the right side, which is linearly spaced between a value £0.5 higher than the low outcome in the lottery and a value £0.5 lower than the high outcome (see Figure A.1 for a sample screen-shot). The basic design consisted of four two-outcome gambles crossed with nine probabilities associated with the high outcome. The two outcomes are (in pounds) (50 - 0), (20 - 0), (15 - 5), (20 - 10). The nine probabilities are .01, .05, .10, .25, .50, .75, .90, 95, and .99. We have in total 36 prospects and hence 36 rounds in Stage 1. The order of the pages is randomised for each subject.

Stage 2

Two types of choice questions are used in Stage 2 to examine the mentioned risk-seeking behaviour and the common consequence effect: the 'mean-preserving' question (a choice between a sure payoff and its mean-preserving spread) and the common consequence effect question (a choice between two lotteries which share a common consequence). After a series of pilots, we chose to have 40 questions in Stage 2 because of budget and time constraints. We decided to use 20 'mean-preserving' questions and 20 common consequence effect questions, with every subject answering the same set of questions with different and randomised orders.[9]

The 'mean-preserving' question takes the following form:

$$A = \begin{cases} x + \frac{(1-p)a}{p}, & p \\ x - a, & 1-p \end{cases} \quad ; \quad B = \begin{cases} x, & 1 \end{cases} \tag{11}$$

where the expected values of A and B are equal. We firstly set $x \in \{5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$, and let a be any positive integer smaller than x. p is set to be a number in $\{0.25, 0.4, 0.5, 0.6, 0.8\}$ for two reasons: First, we pay subjects with real money, so payments with decimal positions less than two are preferred. Therefore, probabilities like 0.33 and 0.66 are not

[9] See Appendix B for questions used in real sessions.

included; Second, we do not consider small p such as 0.01 and 0.05, since small probabilities would make the associated payoffs very large, and we constrain the largest payment in Stage 2 to be £25.[10] This leaves us with 5500 combinations.[11] After discarding questions with excessively high payments and those that are not payable (having more than two decimal points), 159 questions are left, and 20 of them are picked randomly for Stage 2. Only 20 questions are selected out of 5500, yet we argue that the selection is to some extent representative. After cutting off around 5000 combinations, the remaining still cover a wide range of payoffs. For instance, the sure payoff (x) ranges in $\{5,\ 6,\ 7,\ 9,\ 10,\ 11,\ 12,\ 13\}$, and the higher payoff ($x + \frac{(1-p)a}{p}$) in the lottery ranges in $\{6.5,\ 7.0,\ 8.0,\ 8.5,\ 9.0,\ 10.0,\ 11.0,\ 11.5,\ 12.0,\ 13.0,\ 14.5,\ 15.0\}$. Also, a potential drawback of using a comprehensive list of questions is that participants are likely to notice the pattern of the questions even if the order is randomised.

The common consequence effect question takes the following form:

$$C = \begin{cases} h, & p_h \\ z, & p_z \\ 0, & p_0 \end{cases} \quad ; \quad D = \begin{cases} l, & p_h + p_0 \\ z, & p_z \end{cases} \tag{12}$$

where $h > l$ and $z \in [0, l]$. p_h, p_z, and p_0 are the corresponding probabilities for payoffs h, z and 0 respectively. The probability corresponding to l is $p_h + p_0$. This form of pairwise choice can be seen as a generalised form of Tversky and Kahneman (1992)'s common consequence effect example. The consistently observed behaviour is that subjects shift from preferring the riskier option to preferring the safer one when z changes from 0 to l, i.e., when $z = 0$, $C \succ D$, and when $z = l$, $C \prec D$. Also, we believe that this behaviour is more likely to be triggered if h is only slightly larger than l and p_0 is relatively small.[12] Therefore, considering the budget, we set $h \in \{10,\ 11,\ 12,\ 13,\ 14,\ 15\}$, and $l = h - a$ where $a \in \{1,\ 2\}$, and $p_0 \in \{0.01,\ 0.05\}$. Also, we let $p_h \in \{0.1,\ 0.33,\ 0.5,\ 0.75,\ 0.9\}$ and $z \in \{0,\ l/4,\ l/2,\ 3l/4,\ l\}$. This leaves us with 600

[10] Subjects can usually finish Stage 2 in less than 30 minutes, and a possible payment of £25 is sufficiently large.

[11] For any $k \in x$, we have $a \in \{1, 2, ..., k\text{-}2, k\text{-}1\}$. There are five possible probabilities, and $x \in \{5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$. Therefore, the total number of combinations is $5 \times \sum_{k=5}^{15} k(k-1) = 5 \times (\sum_{k=5}^{15} k^2 - \sum_{k=5}^{15} k) = 5 \times (\sum_{k=1}^{15} k^2 - \sum_{k=1}^{4} k^2 - \sum_{k=5}^{15} k) = 5500$.

[12] Allais (1953)'s original example satisfies a similar form, and simple intuition also supports the idea. For instance, if h is far greater than l, it is likely that C is preferred to D when $z = l$, and if p_0 is very large, subjects would preferred D to C when z is 0 as they would not risk a huge chance (p_0) of getting l for a slightly larger outcome h.

combinations, and then we eliminate combinations for which the two theories yield the same predictions conditional on the pilot calibrations ($\alpha = 0.86$,$\gamma = 0.62$, $\theta = 3.28$ and $\delta = 0.67$).[13] 150 questions are left and we randomly choose 20 questions for the remaining part of Stage 2.

Implementation

Our study has a sample size of 48 and subjects are students from the University of Southampton crossing all disciplines, 47% of them being females.[14] The experiment was conducted in the Social Sciences Experimental Laboratory (SSEL) at the University of Southampton using oTree (Chen et al., 2016). The average payment per subject was £16.07 and each session of the experiment lasted around 60 minutes including the payment stage. The payment includes a participation fee of £4, a fixed Stage 1 fee of £4 and the payment for a randomly chosen question of Stage 2 (the payment was determined by a ten-sided die, see details in Appendix A).[15] At the beginning of each session, subjects read the instructions, and the experimenters show them the die which will be used to determine their payoffs at the end of the experiment. Experimenters try to ensure that subjects trust the instructions, and that their payments only depend on their decisions and luck. There is no time constraint in answering the questions, and subjects do not need to wait for others to start the next question or stage. Some subjects may finish before others, and the following statement is displayed on their screens: "We are waiting for everyone to finish. Thank you so much for your patience". After everyone is finished, a random question is selected by the computer for each subject and is displayed on the screen with the subject's decision. The experimenters then go to the subjects one by one to roll the dice and record the payment. Subjects collect their payments and they are free to leave.

[13] In terms of the anomalies which the two theories can explain, CPT and ST still overlap. If the purpose is to 'race' between two theories, examining the overlapping part is not very meaningful. We shall try to focus on questions for which the two theories yield different predictions. Also, please note that the pilot sessions were conducted at the University of Southampton with 42 subjects.

[14] We recruited 49 subjects using ORSEE (Greiner, 2015) at first and exclude one subject because of extreme calibrations. This subject has $\tau = 2.75^{-10}$, $s_{eut} = 6.88^{-9}$, $\gamma = 0$, $s_{cpt} = 0$, and $\delta = 8.04^{-14}$ (see Section 4.1 for the notations). A close look at the data reveals that this subject is extremely risk-averse and they always chose the sure payoff in Stage 1 (for all 288 decisions). This is why τ, γ, and δ are extremely small, and the zero standard deviation of the error term makes it impossible to calculate the out-of-sample likelihoods. Also, it is clear that for such subject, the strategy in Stage 2 would be simply choosing the safer option. It is indeed the case for the subject, and out of 40 questions, they selected the safer option 38 times.

[15] We pay subjects a fixed amount in Stage 1 since a reasonable method (such as the Becker–DeGroot–Marschak method) to incentivise the subjects would increase the experiment time significantly, and the calibrations in our real sessions are comparable to the literature (see Section 4.1).

4 Results

4.1 Calibrations

The maximum likelihood method is used to estimate the parameters and obtain the predicted log-likelihoods, and assumptions need to be made regarding the stochastic nature of the data. Assume that a subject faces n binary choice questions: (L_1,R_1), (L_2, R_2), (L_3, R_3),...,(L_n, R_n), and the preference function is $V(.)$. Then for an arbitrary question i, in the absence of noise, the subject chooses $L_i\,(R_i)$ if $V_{L_i} > V_{R_i}\,(V_{R_i} > V_{L_i})$. With choice error ϵ, the subject prefers $L_i\,(R_i)$ to $R_i\,(L_i)$ only if $V_{L_i} - V_{R_i} + \epsilon > 0\,(V_{R_i} - V_{L_i} + \epsilon > 0)$. Following the Fechner error specification,[16] we assume that the error term ϵ is normally distributed with mean zero and variance s^2. Let $\mathcal{L}_i$ denote the likelihood function of question i, and we have $\mathcal{L}_i = Prob(\epsilon > V_{R_i} - V_{L_i})$. Since $\epsilon \sim \mathcal{N}(0, s^2)$, $\mathcal{L}_i$ equals $1 - \Phi[(V_{R_i} - V_{L_i})/s]$ where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution, and the log-likelihood function is written as follows:

$$\sum_{i=1}^{n} ln(1 - \Phi[(V_{R_i} - V_{L_i})/s]). \tag{13}$$

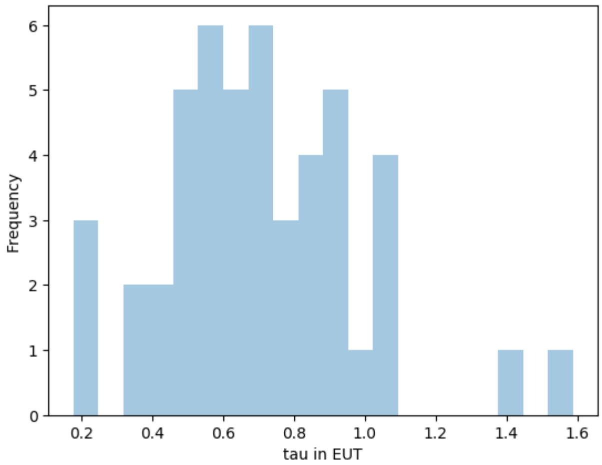
Table 2: Descriptive Summary of Estimates

Theory	Parameter	Mean	Median	s.d.
EUT	τ	0.71	0.69	0.28
	s_{eut}	3.34	1.67	7.22
CPT	α	0.80	0.81	0.18
	γ	0.66	0.64	0.19
	s_{cpt}	2.42	1.61	3.51
ST	θ	29.03	1.23	106.76
	δ	0.67	0.72	0.30
	s_{st}	6.78	5.47	4.70

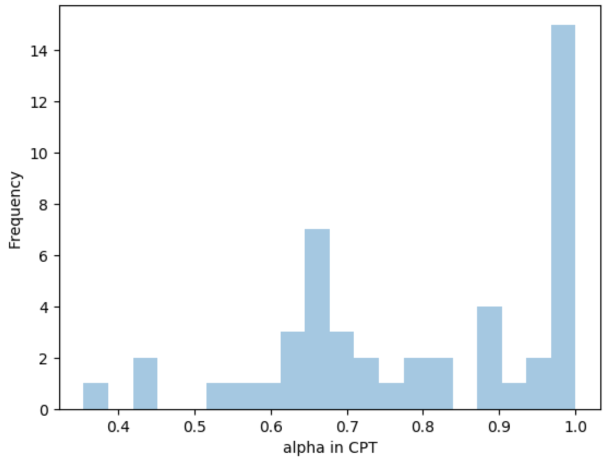
Note: The results exclude one subject for the reasons explained in footnote 14.

[16] We use the Fechner error story in our analysis, since it is relatively simple and most commonly applied in the relevant literature.

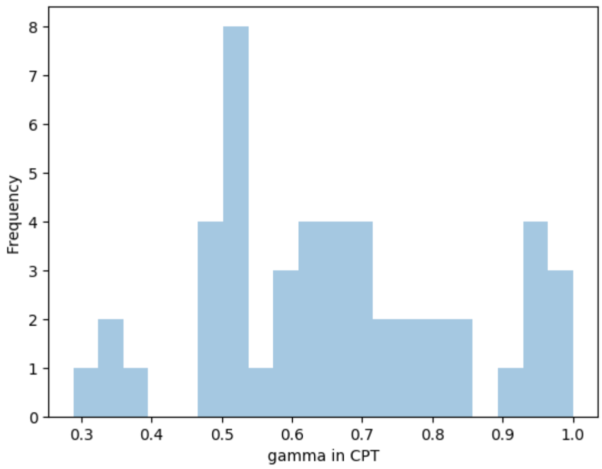
Figure 1: Distributions of estimates



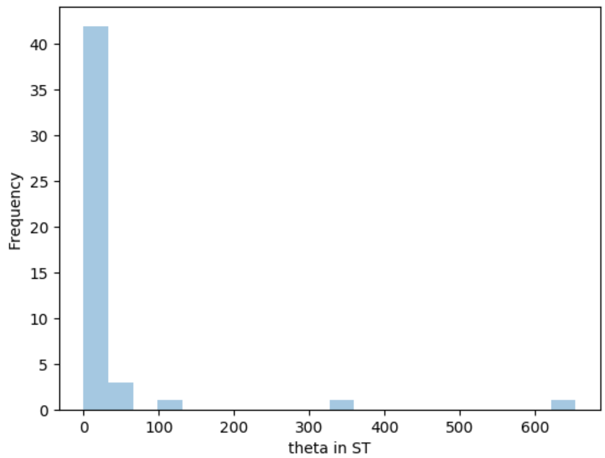
(a)



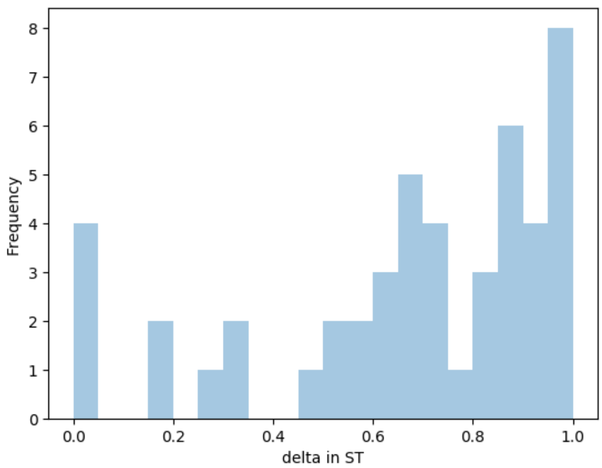
(b)



(c)



(d)



(e)

Observations of Stage 1 are used to pinpoint the functional form. The calibrations are at the individual level, and the summary of the results is presented in Table 2. The calibration result for EUT shows that the majority of subjects are risk-averse (42 out of 48), and for most of them, τ ranges from 0.5 to 0.9 (see Figure 1a). In terms of α and γ in CPT, the distributions of the values are depicted in Figures 1b and 1c. Comparing Figure 1a to Figure 1b, we see that incorporating non-linear probabilities affects the curvature of the value function significantly. Also, the fact that a high fraction of subjects has α near 1 may indicate that the S-shaped probability weighting function is crucial in describing the decision patterns. The calibrations are similar to the pilot sessions in which the medians of α and γ are 0.86 and 0.62 respectively. Besides, they are comparable to the original estimations made by Tversky and Kahneman (1992) (the medians of α and γ are 0.88 and 0.61). Regarding the calibrations of ST, δ is comparable to the pilot session calibrations (a median of 0.67) and Bordalo et al. (2012)'s calibration (0.7). Additionally, it is roughly consistent with Königsheim et al. (2019)'s result which shows that δ is between 0.7 and 0.8. However, according to Table 2, the estimates of θ show significant heterogeneity among individuals. Figure 1d depicts the distribution of θ, which shows that most estimates are smaller than 20. To be more specific, for 39 among 48 subjects, θ is smaller than 10. Considering only those 39 subjects, the mean and median become 1.45 and 0.51, and the s.d. reduces to 1.87. Therefore, our estimates of θ are in fact not decentralised. Outliers with extremely large θ (greater than 400) magnify the standard deviation. We include the outliers into our analysis, since they do not affect the out-of-sample log-likelihoods. However, we acknowledge that our functional form of the salience function is not optimal. Optimising the form of the salience function is the topic of future research. In general, we obtain reasonable estimates of the parameters, and this is important as the result is sensitive to calibrations.

4.2 Descriptive Power

The measure of descriptive power is based on the fitted values of the maximised log-likelihoods in Stage 1. The fitted log-likelihoods are not compared directly since EUT has a different number of parameters relative to CPT and ST. To take into account the different degrees of freedom of the compared models, we apply the Akaike Information Criterion (Akaike, 1973)[17] which takes the following form:

$$AIC = 2k - 2ln(\mathcal{L}), \tag{14}$$

where k denotes the degrees of freedom, and $ln(\mathcal{L})$ represents the log-likelihood function. We firstly compare the mean, the median as well as the 5% and 10% trimmed means of AIC values across all subjects, and we find that CPT, in general, has the best fit and ST outperforms EUT. The results are reported in Table 3 (a smaller AIC value corresponds to a better fit). Besides, Akaike weights (Burnham and Anderson, 2002) are calculated using CPT as the benchmark model.[18] The average Akaike weights for EUT, CPT and ST are 14.97%, 62.53% and 22.5% respectively. This indicates that the probability of CPT being the best fitting model among the candidates is more than 60%, and ST has a slightly better chance than EUT. Also, we analyse the data at the individual level, and find the best fitted model for each subject. We report the percentages of subjects for which a given theory yields the best, second best, and worst fit in Table 4. The '1st' column indicates that for 56.25% of subjects CPT performs best, and the percentages for EUT and ST are 18.75% and 25% (comparable to the Akaike weights). Further, Figure 2 shows the Kernel density estimation of AIC for the three theories and it confirms the notion that CPT outperforms ST, and ST outperforms EUT in terms of descriptive adequacy.

Table 3: Summary of AIC

Theory	Mean	$\text{Mean}_{0.05}$	$\text{Mean}_{0.1}$	Median	s.d.
EUT	202.62	202.67	202.72	200.04	67.42
CPT	165.85⋆	164.56⋆	163.84⋆	153.07⋆	70.21
ST	188.01	190.03	186.95	186.28	77.40

Note: $\text{Mean}_{0.05}$ and $\text{Mean}_{0.1}$ represents the 5% and 10% trimmed means.

⋆ indicates the best performing model in each column.

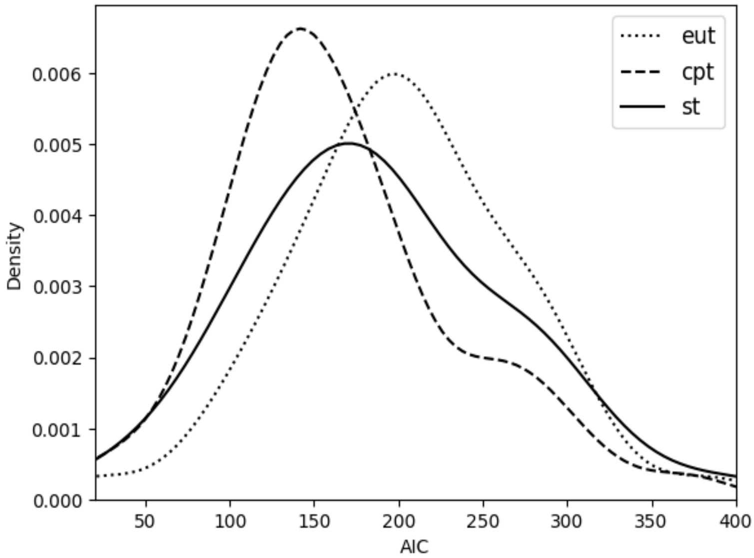
[17] Sugiura (1978) and Hurvich and Tsai (1989) modify the statistic for small-sample studies. The corrected AIC equals to $2k - 2ln(\mathcal{L}) + \frac{2k(k+1)}{n-k-1}$ where n is the number observations, and it punishes the model if there are too many parameters comparing to the sample size. We decide not to implement this correction since for each subject we have $8 \times 36 = 288$ observations, and the degrees of freedom are 1 (for EUT) and 2 (for CPT and ST).

[18] According to Burnham and Anderson (2002), Akaike weights are determined as follows. One wants to select the best model from N candidates (Model 1 to Model N), using Model b as the benchmark model (in practice, Model b is the one which is presumed to be the best). Firstly, one computes $\Delta_i = AIC_i - AIC_b$ for $i \in \{1, 2, 3, ..., N\}$, and AIC_b is the AIC value for model b. Then the Akaike weight $= \frac{\exp(-\Delta_i/2)}{\sum_{n=1}^{N} \exp(-\Delta_n/2)}$. The interpretation is straightforward: Akaike weights indicate the probability that a model is the best among the whole set of candidates.

Table 4: Ranking based on AIC

Theory	1st	2nd	3rd
EUT	18.75%	22.92%	58.33%
CPT	56.25%	37.50%	6.25%
ST	25.00%	39.58%	35.42%

Figure 2: Kernel density estimation of AIC



4.3 Predictive Power

Combined with the estimated model, the observations of each subject in Stage 2 are put back into Equation 13 to calculate the predicted log-likelihood which is used to measure the predictive power. The predicted log-likelihoods are analysed without any modification related to different degrees of freedom of theories, since overfitted models have disadvantages when comparing out-of-sample log-likelihoods (Hey et al., 2010). Table 5 reports the major statistics about predicted log-likelihoods. A similar pattern to descriptive power is found: CPT dominates among the three theories, and ST outperforms EUT.[19] This time, however, CPT outperforms ST by a very insignificant margin (-25.41 vs -25.98). Similarly, the Kernel density estimation (Figure 3) shows that the predicted log-likelihoods of EUT are centralised between roughly -30 and -25, and the predicted log-likelihoods of CPT and ST are decentralised with

[19]On the one hand, all models outperform a random choice mechanism (on average). The predicted log-likelihood of a 'coin-tosser' (an agent who flips a coin every time he chooses between two alternatives) making 40 decisions is $ln(0.5) \times 40 \approx -27.73$. On the other hand, behaviours vary from individual to individual. In fact, there are around 29.2% of subjects who behave more similarly to a 'coin-tosser', rather than to an EUT agent. The percentages for CPT and ST are 12.2% and 22.9%.

similar distributions (fat tail on the right side). Further, in Table 6, we report the percentage of subjects for whom each theory is dominating. Specifically, CPT wins for the majority of the subjects (60.42%), while for roughly half of the subjects (56.25%), ST is the second best, and EUT is 'the worst performer' for 58.33% of the subjects.

Further, Figure 4 shows the scatter plots of predicted log-likelihood against AIC for the three theories, and it demonstrates a negative correlation between the two statistics, i.e., positive correlation between descriptive and predictive power. The implication that theories with better descriptive ability are associated with higher predictive ability, in line with previous experiments, such as Hey et al. (2010), Hey and Pace (2014) and Georgalos (2019).

Table 5: Summary of predicted log-likelihoods

Theory	Mean	$\text{Mean}_{0.05}$	$\text{Mean}_{0.1}$	Median	s.d.
EUT	-27.20	-27.34	-27.35	-27.40	1.51
CPT	-25.41⋆	-25.11⋆	-25.27⋆	-26.12⋆	5.82
ST	-25.98	-26.01	-26.09	-26.63	2.81

Note: $\text{Mean}_{0.05}$ and $\text{Mean}_{0.1}$ represents the 5% and 10% trimmed means.
⋆ indicates the best performing model in each column.

Table 6: Ranking based on predicted log-likelihoods

Theory	1st	2nd	3rd
EUT	18.75%	22.92%	58.33%
CPT	60.42%	20.83%	18.75%
ST	20.83%	56.25%	22.92%

4.4 Additional Analysis

Linear Utility vs Non-linear Utility: How Does It Matter?

The above results show that ST achieves predictive power similar to CPT, and it does not require a non-linear value function. In the framework of ST, most shifts of risk preferences can be explained by the 'ordering' and the 'diminishing sensitivity' properties of the salience function, as well as the function's convexity (Bordalo et al., 2012, p. 1278 -1279),[20] and the shape

[20]According to Bordalo et al. (2012), a salience function is convex if diminishing sensitivity becomes weaker as the overall payoff gets higher.

Figure 3: Kernel density estimation of predicted log-likelihoods

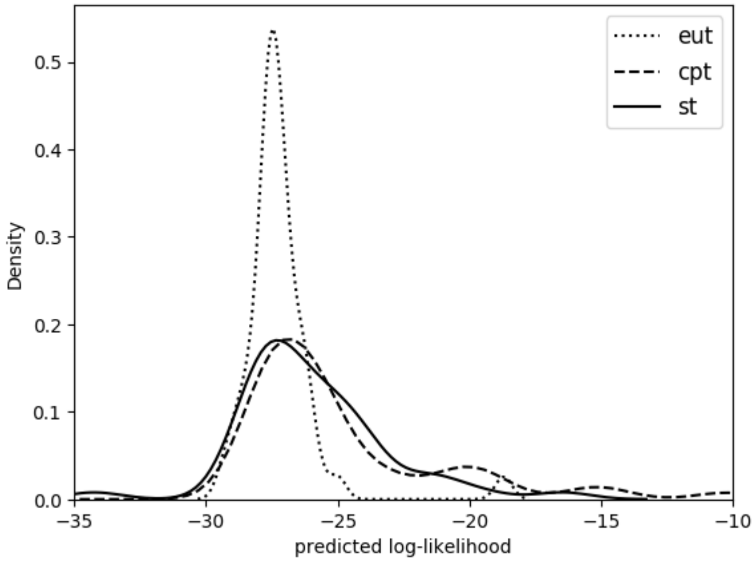
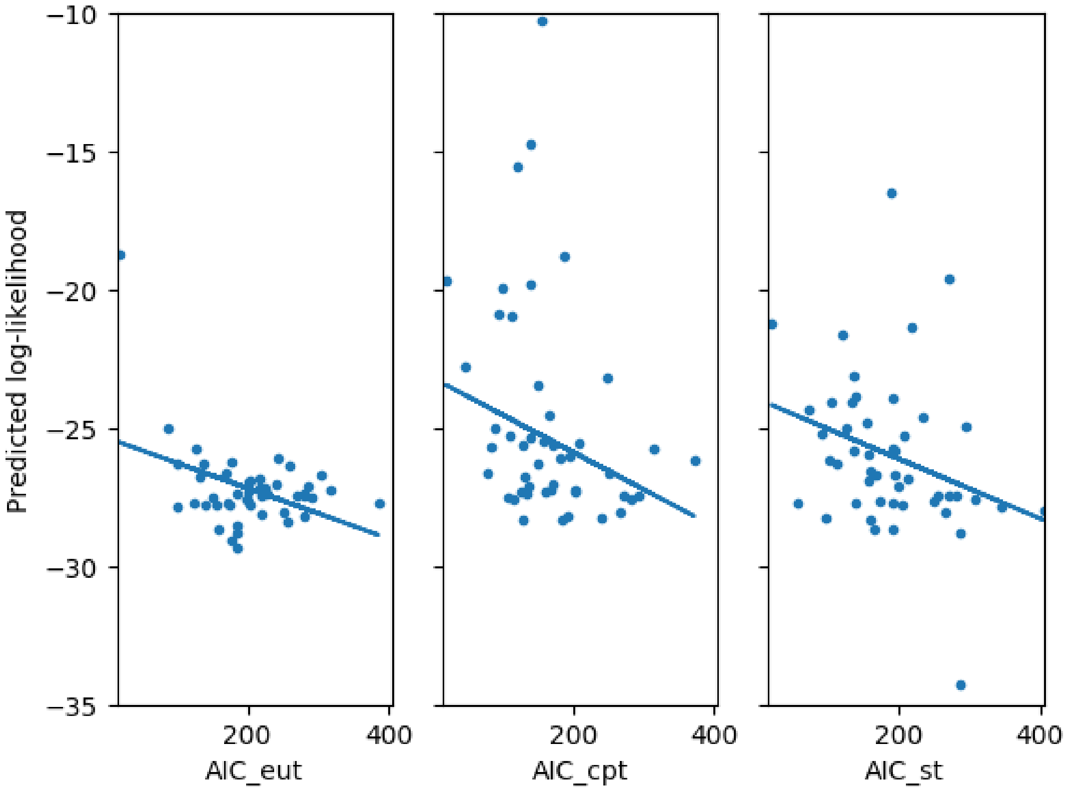


Figure 4: Scatter of predicted log-likelihood against AIC



of the value function does not play a crucial role. However, the curvature of the value function is important for CPT. We examine how linear/non-linear utility affect the performance of both theories by introducing one variation of each. One is CPT with a linear value function (henceforth denoted as LCPT), and the other one is ST with a non-linear value function (henceforth denoted as NST). The parametric representation of LCPT is identical to CPT, except that the value function is linear, i.e., $v(x) = x$. The parametric representation of NST is identical to

ST, except that the value function is non-linear, i.e., $v(x) = x^{\tau'}$, where $\tau' > 0$. For LCPT, the weighting function is $\frac{p^{\gamma'}}{(p^{\gamma'}+(1-p)^{\gamma'})^{1/\gamma'}}$, where $0 < \gamma' < 1$. For NST, the salience function is: $\frac{|x_i^s - x_j^s|+\theta'}{|x_i^s|+|x_j^s|+e\theta'}$, where $\theta' > 0$, and the local thinking parameter is $\delta' \in (0, 1]$. The summary of the estimated parameters and the distributions of the estimates are presented in Appendix C.

We report the average level of AIC and predicted log-likelihoods of LCPT and NST in Table 7. Comparing the result in Table 7 with the results in Table 3 and Table 5, we see that the statistics for LCPT and NST roughly lie between CPT and ST. The pattern is clear for AIC, but not for predicted log-likelihoods, as the gap between CPT and ST is small in the first place. We shall take a close look at the Kernel density estimations in Figure 5. Despite the fact that the differences are small, it is unambiguous that, comparing to CPT, the descriptive and predictive performance of LCPT is closer to ST and NST does not perform better than ST. Also, it is worth noticing that in terms of applications, the calibration for LCPT is stabler than ST, and ST needs one extra degree of freedom. We believe that further work should focus on the parametric form of the salience function, since assuming non-linear utility does not improve performance.

Table 7: Summary of AIC and predicted log-likelihoods (LCPT and NST)

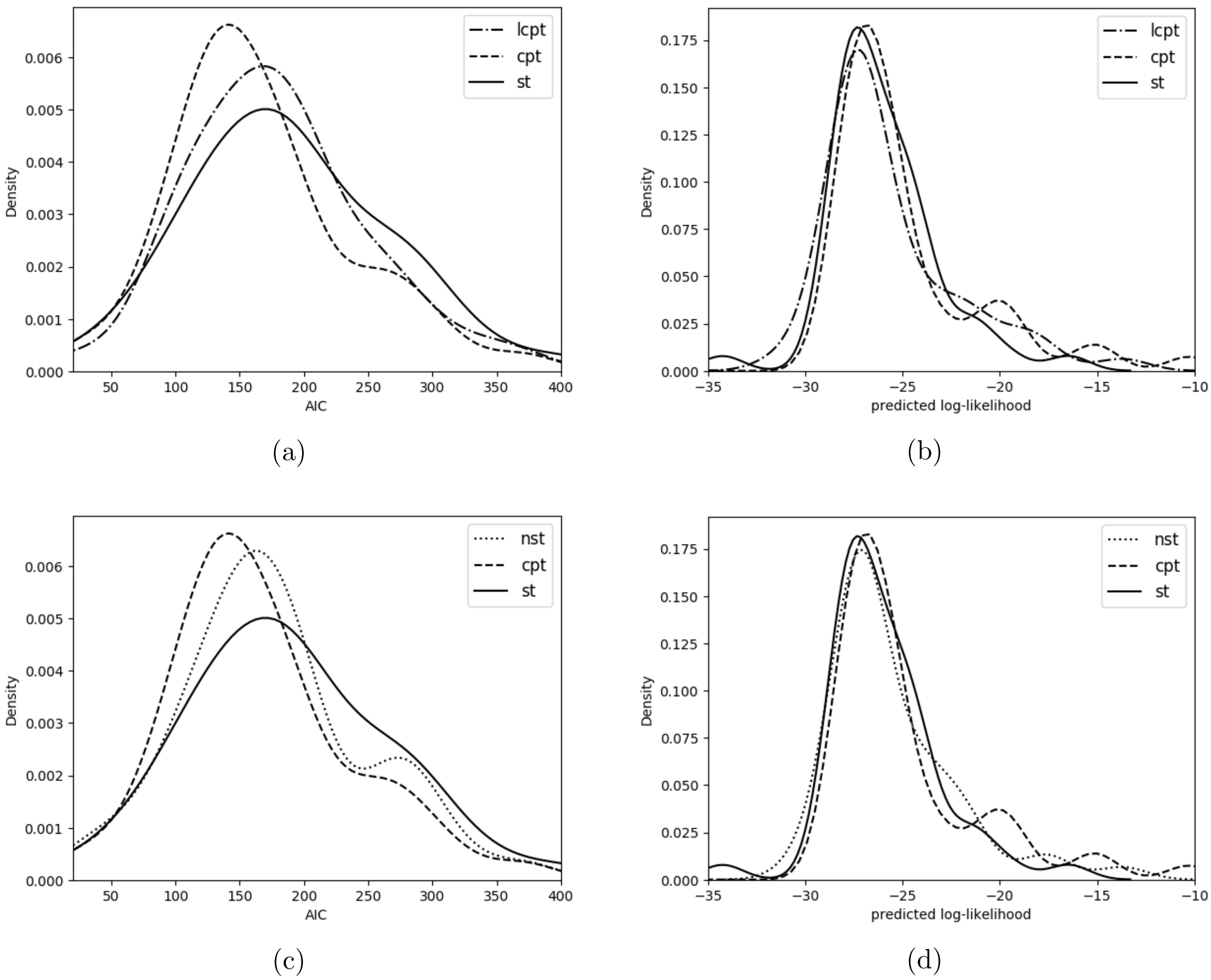
		Mean	$\text{Mean}_{0.05}$	$\text{Mean}_{0.1}$	Median	s.d.
LCPT	AIC	178.65	177.93	176.66	168.82	68.87
	predicted	-25.62	-25.76	-25.85	-26.84	3.47
NST	AIC	176.07	175.20	174.85	163.57	72.10
	predicted	-25.54	-25.69	-25.80	-26.64	3.20

Note: $\text{Mean}_{0.05}$ and $\text{Mean}_{0.1}$ represents the 5% and 10% trimmed means.

The Local Thinking Parameter

The local thinking parameter is δ in Equation 6, and it measures how much a local thinker differs from a rational economic decision maker (when $\delta = 1$, the decision maker does not distort the decision weights and ST reduces to expected value with the linear utility function). Königsheim et al. (2019) focus on this parameter and three of their main results are: i) the parameter is between 0.7 and 0.8; ii) the parameter does not change much when non-linear utility is assumed; iii) the parameter is not stable, in the sense that it is smaller if the choice

Figure 5: Kernel density estimations of AIC and predicted log-likelihood (LCPT and NST)



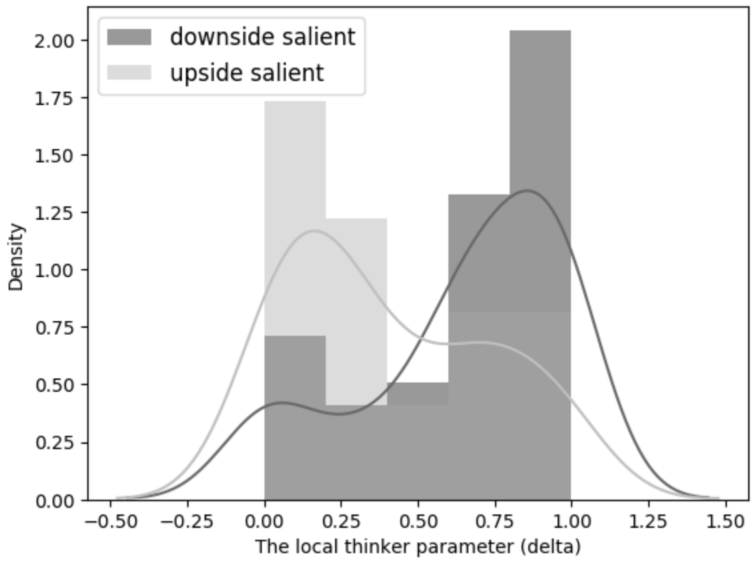
question has a salient downside.[21] Our estimates of δ and δ' are about 0.7 which are in line with their first and second results (see Section 4.1 and Appendix C). However, further analysis reveals an exactly opposite result than their third result, i.e., we find heterogeneity, but a salient downside corresponds to significantly larger δ.

The design of Stage 1 provides a natural setting to examine this possible heterogeneity in δ. There are eight questions in each round of Stage 1, and four of them have a salient downside and the other four have a salient upside. Taking the sample question in Figure A.1 as an example, apparently, the upper four options in OPTION B column make the downside of OPTION A (receiving 0 pounds) more salient, while the bottom four options make the upside (receiving 20 pounds) more salient. We apply MLE to the observations of downside

[21] A choice question is downside salient, if the low outcomes of the relative riskier lottery are in the most salient states. Therefore, when facing such question, a local thinker is more likely to focus on the downside of the riskier option and prefers the safer option.

and upside questions separately, and the estimated medians of local thinking parameters are 0.72 for questions with salient downside and 0.33 for questions with salient upside. In terms of statistical significance, the p-value of Wilcoxon rank-sum test is 0.0009, and the alternative is that values for the salient downside setting are more likely to be larger than the values in the upside setting. The distributions of the values in Figure 6 confirm this. We suggest that considering the heterogeneity of δ is a promising future direction for enhancing ST's predictive power.[22]

Figure 6: Distributions of the local thinking parameter: downside salient vs. upside salient



5 Conclusion

We empirically test Bordalo et al. (2012)'s ST in settings where the majority of subjects exhibit unstable risk attitudes. Firstly, we provide a formal calibration of ST, which has only been done by Königsheim et al. (2019). We find that the local thinking parameter is about 0.7, and the result is consistent with Königsheim et al. (2019)'s. Then, we compare ST and CPT on the basis of the in-sample and out-of-sample log-likelihoods. Two types of binary choice

[22] A simple exercise using the observations in Stage 2 shows a slight increase in predictive power if we consider the heterogeneity of δ. Questions in Stage 2 can be categorised according to the 'directions of salience'. In terms of the 'mean-preserving' questions in Equation 11, $p \geq \frac{1}{2}$ indicates a salient downside and $p < \frac{1}{2}$ indicates a salient upside. In terms of the common consequence effect in Equation 12, a sufficiently small z indicates a salient downside (we consider z small if $z \in \{0, l/4\}$). Out of 40 questions in Stage 2, 25 of them are downside salient and the remaining questions are upside salient. If the predicted log-likelihoods of questions with different salience directions are calculated according to the corresponding calibrations, the mean of predicted log-likelihoods becomes -25.64 which is larger than the -25.98 of Table 5.

questions, which reveal the instability of subjects' risk preferences are used: 'mean-preserving' questions and common consequence effect questions. We find that CPT outperforms ST in terms of both descriptive and predictive power. However, the gaps are small, especially for predictive power. After a further investigation on the linearity of the value function and the heterogeneity of the local thinking parameter, we envision two possible pathways for improving ST for pure application purposes: developing a new specification of the salience function and analysing upside and downside salient settings with different local thinking parameters.

We shall conclude by an assessment of ST and its prospect as a new alternative to CPT. ST presents reasonable descriptive and predictive power, in comparison to the highly popular CPT. In our view, salience, in particular, the notion that people focus their attention on the most salient aspect of the world, plays a indisputable role in choice under risk. Salience can also be important for a variety of other economics situations, such as taxation (Chetty et al., 2009), asset pricing (Bordalo et al., 2013a), consumer behaviour (Bordalo et al., 2013b) and judicial decisions (Bordalo et al., 2015) etc. We anticipate that future empirical studies shall study salience in a broader domain of judgement setting.

References

Akaike, H. (1973). Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, 60(2):255–265.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica*, 21(4):503–546.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics*, 127(3):1243–1285.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013a). Salience and asset prices. *The American Economic Review*, 103(3):623–628.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013b). Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2015). Salience theory of judicial decisions. *The Journal of Legal Studies*, 44(S1):S7–S33.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference*. Springer.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–77.

Dertwinkel-Kalt, M. and Köster, M. (2019). Salience and skewness preferences. *Journal of the European Economic Association*. https://doi.org/10.1093/jeea/jvz035.

Frydman, C. and Mormann, M. M. (2018). The role of salience in choice under risk: An experimental investigation. *Available at SSRN 2778822*.

Georgalos, K. (2019). An experimental test of the predictive power of dynamic ambiguity models. *Journal of Risk and Uncertainty*. forthcoming.

Gonzalez, R. and Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, 38(1):129–166.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.

Hey, J. D., Lotito, G., and Maffioletti, A. (2010). The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity. *Journal of Risk and Uncertainty*, 41(2):81–111.

Hey, J. D. and Pace, N. (2014). The explanatory and predictive power of non two-stage-probability theories of decision making under ambiguity. *Journal of Risk and Uncertainty*, 49(1):1–29.

Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, pages 297–307.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pages 263–291.

Kahneman, D. and Tversky, A. (1981). The simulation heuristic. Technical report, Stanford University.

Königsheim, C., Lukas, M., and Nöth, M. (2019). Salience theory: Calibration and heterogeneity in probability distortion. *Journal of Economic Behavior & Organization*, 157:477–495.

Kontek, K. (2016). A critical note on salience theory of choice under risk. *Economics Letters*, 149:168–171.

Kothiyal, A., Spinu, V., and Wakker, P. P. (2014). An experimental test of prospect theory for predicting choice under ambiguity. *Journal of Risk and Uncertainty*, 48(1):1–17.

Lichtenstein, S. and Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 89(1):46.

Nielsen, C. S., Sebald, A. C., and Sørensen, P. N. (2018). Testing for salience effects in choices under risk. Technical report, Working Paper, University of Copenhagen.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343.

Sugiura, N. (1978). Further analysts of the data by akaike's information criterion and the finite corrections. *Communications in Statistics-Theory and Methods*, 7(1):13–26.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Wu, G. and Markle, A. B. (2008). An empirical test of gain-loss separability in prospect theory. *Management Science*, 54(7):1322–1335.

Appendices

A Instruction

Introduction

Welcome and thank you for participating in today's experiment. It is important that you do not talk, or in any other way try to communicate during the study. You cannot use your phone during the study. If you have any questions, please just raise your hand and wait for the assistance.

During this experiment you will earn money. How much you earn depends on your decisions and luck. The money will be paid to you, in cash, at the end of the experiment.

Your participation in the experiment and any information about you will be kept anonymised and confidential. Your receipt of payment and consent form are the only places on which your name will appear. This information will be kept confidential in the manner described in the consent form.

The experiment has two stages. Stage 1 consists of 36 periods. In each period, you will make choices between a lottery and a series of certain values. In Stage 2, you will make 30 choices. There will be a short questionnaire at the end.

Stage 1

The first stage will have 36 periods. In each period, the screen displays a lottery (Option A) on the left and displays a descending series of eight sure outcomes (Option B) on the right side. Every period has the same general form.

An example of the screen for a period is given in Figure A.1. As you can see, for each decision, you must choose between Option A and Option B. You may choose Option A for some decisions and Option B for others, and you may change your decisions and make them in any

Figure A.1: Stage 1 sample screen

For each row of the following table,please decide whether you prefer **OPTION A(a risk prospect)** or **OPTION B(a certain payoff)**.

OPTION A(in pounds)	**Your Choices**	**OPTION B(in pounds)**
Receiving **20.0** with probability **0.9** Receiving **0.0** with probability **0.1**	○A ○B	Receiving **19.5** for sure
	○A ○B	Receiving **17.0** for sure
	○A ○B	Receiving **14.0** for sure
	○A ○B	Receiving **11.5** for sure
	○A ○B	Receiving **8.5** for sure
	○A ○B	Receiving **6.0** for sure
	○A ○B	Receiving **3.0** for sure
	○A ○B	Receiving **0.5** for sure

Submit

order. Once you have made all of your decisions, press the Submit button and you will be taken to the next period. Note that after you have pressed the submit button, you will no longer be able to change your decisions.

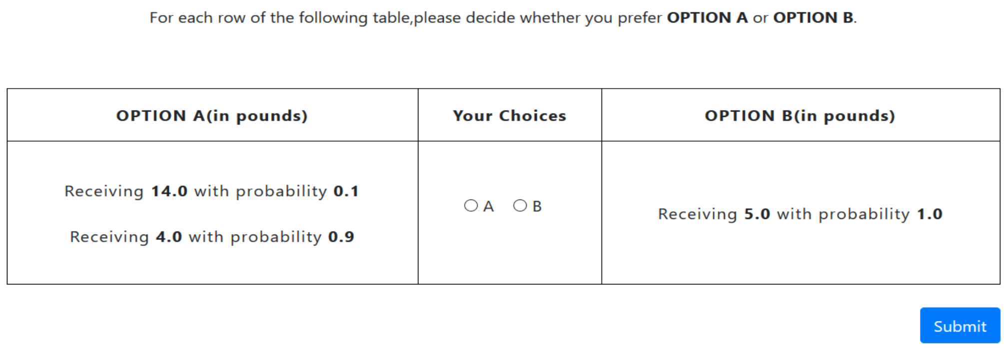
Stage 2

In this stage, you will need to make 30 choices. An example of the screen for a question is given in Figure A.2. As you can see, you simply choose between Option A and Option B, and press the Submit button to go to the next question. Similar to Stage 1, you cannot change your decision after you have pressed the submit button.

Your payment

- You will receive a participation fee of £4 regardless of your decisions.
- You can get an additional Stage 1 fee equal to £4. Note that, if you complete Stage 1, however leave the experiment before Stage 2 is finished, you will only earn the participa-

Figure A.2: Stage 2 sample screen


For each row of the following table,please decide whether you prefer **OPTION A** or **OPTION B**.

OPTION A(in pounds)	Your Choices	OPTION B(in pounds)
Receiving **14.0** with probability **0.1** Receiving **4.0** with probability **0.9**	○A ○B	Receiving **5.0** with probability **1.0**

Submit

tion fee of £4.

- After you complete Stage 2, the computer will randomly choose a question from Stage 2 and you will be paid according to this question. Depending on the random question and the decision of yours, there are two possibilities:

  1 If the Option you chose in the random question is paying an amount for sure, then that amount will be your payment in Stage 2.

  2 If the Option you chose in the random question is a lottery, you will roll a ten-sided die to determine your payoff. For example, suppose that that Option is as follows:

$$\begin{cases} \text{Receiving £25 with probability 0.1} \\ \text{Receiving £5 with probability 0.9} \end{cases}$$

  In this case, you will roll the ten-sided die once. If a 1 comes up, then you will receive £25 , while if a 2, 3, 4, 5, 6, 7, 8 ,9 or 0, comes up then you will receive £5. As a further example, suppose that the probabilities in that Option have two decimal digits:

$$\begin{cases} \text{Receiving £20 with probability 0.33} \\ \text{Receiving £15 with probability 0.66} \\ \text{Receiving £0 with probability 0.01} \end{cases}$$

  In this case, you will roll the die twice. These two rolls will correspond to a number from 00 to 99. For instance, if you roll 0, 4, it corresponds to 04; if you roll 4, 0, it corresponds to 40. In this example, if the corresponding number you roll is 01, 02, . . . , or 33, then you will receive £20. If your roll is 34, 35, . . . ,or 99, then you will receive £15. If you roll 00, then you will receive £0.

- Your payment = Participant fee (£4) + Payment in Stage 1 (£4) + Payment in Stage 2 (£0 - £15)

If you have any questions, please raise your hand now, otherwise we will begin with the experiment.

B Questions in Stage 2

§ indicates 'mean-preserving' questions. † indicates upside salient questions. Note that in real sessions, we randomised the display order of the questions for each subject. However, the left-right juxtaposition remained the same as a consequence of software limitations.

1. $L = (6, 1), \quad R = (12, 0.4; 2, 0.6)$§†
2. $L = (6, 0.5; 4; 0.5), \quad R = (5, 1)$§
3. $L = (7, 1), \quad R = (8.5, 0.4; 6, 0.6)$§†
4. $L = (7, 1), \quad R = (9, 0.6; 4, 0.4)$§
5. $L = (7, 1), \quad R = (12, 0.5; 2, 0.5)$§
6. $L = (8, 1), \quad R = (12.5, 0.4; 5, 0.6)$§†
7. $L = (8, 0.5; 6, 0.5), \quad R = (7, 1)$§
8. $L = (9, 1), \quad R = (15, 0.25; 7, 0.75)$§†
9. $L = (9, 0.25; 5, 0.75), \quad R = (6, 1)$§†
10. $L = (9, 0.5; 5, 0.5), \quad R = (7, 1)$§
11. $L = (9, 0.76; 6.5, 0.24), \quad R = (10, 0.75; 6.5, 0.24; 0, 0.01)$
12. $L = (10, 1), \quad R = (11, 0.8; 6, 0.2)$§
13. $L = (10, 1), \quad R = (11.5, 0.4; 9, 0.6)$§†
14. $L = (10, 0.11; 2.5, 0.89), \quad R = (12, 0.1; 2.5, 0.89; 0, 0.01)$†
15. $L = (10, 0.55; 2.5, 0.45), \quad R = (12, 0.5; 2.5, 0.45; 0, 0.05)$†
16. $L = (10, 0.11; 5, 0.89), \quad R = (11, 0.1; 5, 0.89; 0, 0.01)$
17. $L = (10, 0.76; 5, 0.24), \quad R = (11, 0.75; 5, 0.24; 0, 0.01)$

18. $L = (10, 0.75; 6, 0.2; 0, 0.05)$, $R = (8, 0.8; 6, 0.2)$

19. $L = (10, 0.5; 6.5, 0.49; 0, 0.01)$, $R = (9, 0.51; 6.5, 0.49)$

20. $L = (10, 0.34; 7.5, 0.66)$, $R = (11, 0.33; 7.5, 0.66; 0, 0.01)$

21. $L = (11, 1)$, $R = (12, 0.5; 10, 0.5)$§

22. $L = (12, 0.11; 0, 0.89)$, $R = (13, 0.1; 0, 0.9)$†

23. $L = (12, 1)$, $R = (15, 0.8; 0, 0.2)$§

24. $L = (12, 0.4; 7, 0.6)$, $R = (9, 1)$§

25. $L = (12, 0.8; 7, 0.2)$, $R = (11, 1)$§

26. $L = (13, 0.9; 0, 0.1)$, $R = (12, 0.95; 0, 0.05)$†

27. $L = (13, 0.5; 2.5, 0.45; 0, 0.05)$, $R = (11, 0.55; 2.5, 0.45)$†

28. $L = (13, 0.11; 3, 0.89)$, $R = (15, 0.1; 3, 0.89; 0, 0.01)$†

29. $L = (13, 0.11; 6.5, 0.89)$, $R = (15, 0.1; 6.5, 0.89; 0, 0.01)$

30. $L = (13, 0.33; 8, 0.66; 0, 0.01)$, $R = (11, 0.34; 8, 0.66)$

31. $L = (13, 0.51; 9.5, 0.49)$, $R = (14, 0.5; 9.5, 0.49; 0, 0.01)$

32. $L = (14, 0.9; 0, 0.1)$, $R = (13, 0.95; 0, 0.05)$†

33. $L = (14, 0.75; 6.5, 0.24; 0, 0.01)$, $R = (13, 0.76; 6.5, 0.24)$

34. $L = (14, 0.33; 9.5, 0.66; 0, 0.01)$, $R = (13, 0.34; 9.5, 0.66)$

35. $L = (14, 0.75; 9.5, 0.24; 0, 0.01)$, $R = (13, 0.76; 9.5, 0.24)$

36. $L = (14.5, 0.8; 2, 0.2)$, $R = (12, 1)$§

37. $L = (15, 0.9; 3.5, 0.09; 0, 0.01)$, $R = (14, 0.91; 3.5, 0.09)$†

38. $L = (15, 0.1; 5, 0.9)$, $R = (6, 1)$§†

39. $L = (15, 0.25; 11, 0.75)$, $R = (12, 1)$§

40. $L = (15, 0.5; 13, 0.5)$, $R = (14, 1)$§

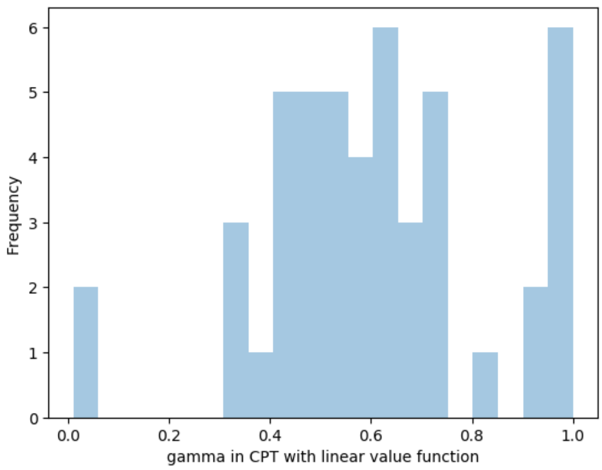
C Additional Tables and Graphs

Table C.1: Descriptive Summary of Estimates (LCPT and NST)

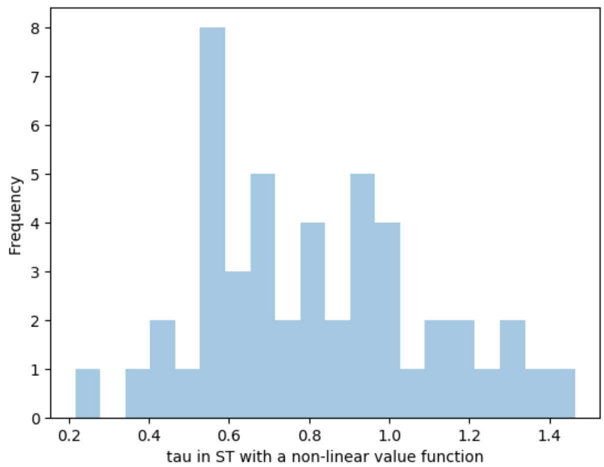
Theory	Parameter	Mean	Median	s.d.
LCPT	γ'	0.60	0.58	0.22
	s_{lcpt}	4.53	3.32	3.94
NST	τ'	0.82	0.81	0.28
	θ'	72.94	1.97	202.62
	δ'	0.66	0.77	0.32
	s_{nst}	3.98	2.26	4.60

Note: The results exclude one subject for the reasons explained in footnote 14.

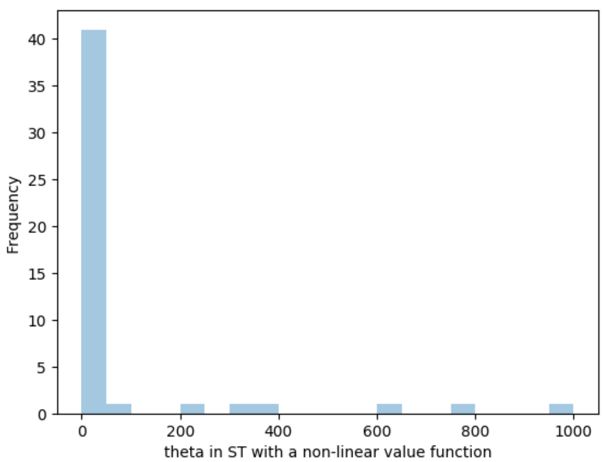
Figure C.1: Distributions of estimates for LCPT and NST



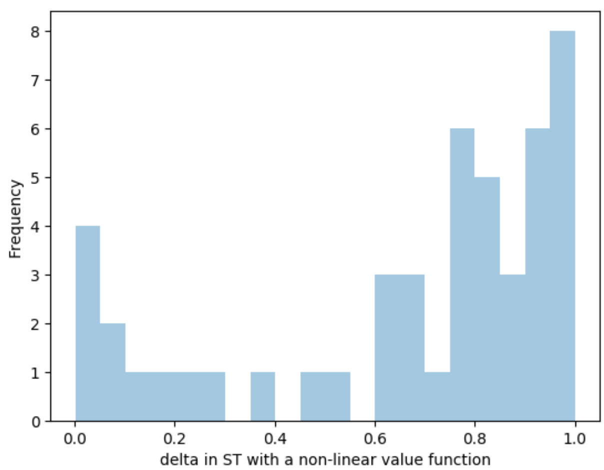
(a)



(b)



(c)



(d)