國立臺灣大學理學院心理學系

碩士論文

Department of Psychology

College of Science

National Taiwan University

Master's Thesis

二分法與適測性心理物理方法於偏好研究中無異點估計之比較

Comparison of Bisection-Based and Adaptive Psychophysical Methods for Eliciting Indifference Points in Preference Research

張瑋宸

Wei-Chen Chang

指導教授: 徐永豐 博士

Advisor: Yung-Fong Hsu Ph.D.

中華民國 114 年 7 月

July 2025

doi:10.6342/NTU202501601

Acknowledgements

大學和碩士班在台大共六年，現在要暫告一段落。過去學習經常是走馬看花地修課及決定下一步，有時也會懷疑這只是無意義地繞遠路。不過隨著碩一下決定題目後，隨著研究進展竟將過去看似無關的經歷串連起來，才意識到過去的學習不是白費。六年多來受到許多師長的指導，以及和不少前後輩交流，在此感謝這些學習路上的人事物。

首先要先感謝指導教授徐永豐老師，從大一的心統課程開始，雖然內容聽的似懂非懂，但讓我對統計方法學有初步的認識、也引發自己想深入了解統計學的興趣，否則過去我總認為數學只限於教科書內的解題過程。大學中間幾年都沒有加入實驗室，直到碩班時才重新找上老師當指導教授。徐老師在指導上給予我很大的自由度，於此同時也給予我許多協助，自決定題目到研究過程以及論文撰寫的過程都是。老師勤於信件來往，也隨時願意和我討論研究上的各種問題，除了提供許多實際的建議之外，有時還會鼓勵沒有自信的自己。在老師的指導下我受益良多。

也感謝實驗室的成員們。鄭澈和吳紘兩位學長不吝於在 meeting 時提供回饋，無論是在研究內容或者是報告的方式上提供很多值得學習的地方，尤其是吳紘學長總是能明確地指出許多我想法的缺陷以及提供建議，在研究方向上提供很大的協助。亦感謝實驗室的同伴們：庭妤、渭倫、筠琪，每週在南館一樓開的小組會議聊天總是很歡樂。

此外要謝謝 EDP 計畫的董建華學長以及多年來所有參與的成員—有些可能認識見過面有些則無—雖然自己只在大二參與短短一學期的時間，但若沒有計畫的

doi:10.6342/NTU202501601

初步構想和過去收集的資料，這篇論文的發想和大半內容都不會存在。

還要感謝兩位口試委員陳暐老師與林逸軒老師，大四在陳暐老師的實驗室接觸到數種實驗經濟學的測量偏好與主觀信念的方法，成了這個研究的靈感來源之一。而林逸軒老師則是願意讓一個外系學生在學期過了快一半時才參加課程旁聽、還在下課時問問題，甚至願意接受口試的邀請，過去自己讀論文看不懂的地方也因為選擇理論的課程有了更清楚的理解。兩位委員亦在口試時提供很多意見和建議，替論文提供了新的想法。

最後感謝家人在生活上的支持，六年也是很長的時間，有你們百分之百的信任讓我可以專注於學業上是無比的幸運；亦感謝至今曾陪伴過的朋友夥伴們，無論系上系外，就不一一列舉，大學到研究生生活能有這些陪伴、情感支持，我才能走到現在。

未來漫漫長路還不知道要如何，但希望能夠保持熱誠前進。

doi:10.6342/NTU202501601

摘要

二分法 (the bisection method) 被應用於所謂的權衡法 (trade-off paradigm, Abdellaoui, 2000; Wakker and Deneffe, 1996)，以測量決策理論中評估效用的無異點。然而其精確度可能受到兩項因素的影響：一是參與者選擇反應中的隨機性，二是二分演算法本身的邊界設定規則。本研究旨在檢驗二分法 (包含其簡化版本 SimpBisection) 之有效性，並探討其他適測性心理物理學方法，如 ASA、PEST 與 MOBS，作為估計無異點的可能替代方法。本研究進行了兩項模擬實驗，採用 Abdellaoui et al. (2016) 中用以測量價值函數的實驗設計，比較各種無異點估計方法的表現。模擬結果顯示，在相同終止準則下，所有方法的估計大致不偏。然而，若邊界設定不當，二分法可能會產生有偏估計。在測試的方法中，ASA 有最高的效率性，但通常需要較多次的迭代數。當迭代次數有限 (但非極少) 時，對選擇行為較為確定的參與者而言，採用固定初始邊界的二分法與 SimpBisection 均能展現良好的效率性。相對地，對於選擇行為較隨機性的參與者，ASA 與 SimpBisection 則為較佳的選擇。鑑於實際實驗常見參與者異質性及限制迭代次數的情況，SimpBisection 可視為各方法間的良好折衷方案。

關鍵字：二分法、無異點、損失趨避、展望理論、模擬研究、隨機逼近法、價值函數

doi:10.6342/NTU202501601

doi:10.6342/NTU202501601

Abstract

The bisection method has been implemented within the so-called trade-off paradigm (Abdellaoui, 2000; Wakker and Deneffe, 1996) to elicit indifference points for utility assessment in decision theory. However, its precision may be affected by two factors: response randomness in participants' choices and the boundary determination rules inherent in the bisection algorithm. This study aims to evaluate the validity of the bisection method — including a simplified version, SimpBisection — and to explore adaptive psychophysical methods such as ASA, PEST, and MOBS as potential alternatives for eliciting indifference points. Two simulation studies were conducted to compare the performance of these elicitation methods under the experimental design of Abdellaoui et al. (2016) for measuring the value function. The simulation results show that, under the same stopping criterion, all methods are largely unbiased. However, the bisection method can produce biased estimates if the boundary settings are poorly chosen. Among the tested methods, ASA demonstrates the highest efficiency, although it typically requires more iterations to complete the procedure. When the number of iterations is limited (but not too small), both the bisection method (with fixed initial boundaries) and SimpBisection perform efficiently for participants exhibiting more deterministic choice behavior. For participants

doi:10.6342/NTU202501601

with greater choice randomness, ASA and SimpBisection offer better alternatives. Given the prevalence of participant heterogeneity and practical constraints on iteration counts in real experiments, SimpBisection appears to be a reasonable compromise.

Keywords: bisection, indifference point, loss aversion, prospect theory, simulation, stochastic approximation, value function

doi:10.6342/NTU202501601

Table of Contents

doi:10.6342/NTU202501601

doi:10.6342/NTU202501601

List of Figures

doi:10.6342/NTU202501601

doi:10.6342/NTU202501601

List of Tables

doi:10.6342/NTU202501601

doi:10.6342/NTU202501601

Chapter 1 Introduction

In the study of preference, a common procedure often involves judging between uncertain or risky alternatives, referred to as prospects (or called gambles, lotteries). For instance, in the certainty equivalent method, the decision maker is asked to determine a sure amount that is equivalent to a risky prospect. In the probability equivalent method, the probability of outcomes in a risky prospect is adjusted by the decision maker to match the equivalence to a sure amount. And in the trade-off method, one outcome in one of the prospects is adjusted by the decision maker to establish equivalence between the two prospect options (Wakker and Deneffe, 1996; Wakker, 2010).

Across these procedures, in each trial one target property, such as the size of an outcome or the probability of an event, may vary in one of the two options, and requires the decision maker to identify the specific quantity that makes the two options indifferent. To obtain this so-called "indifference point," a straightforward approach is to directly ask participants to report it (i.e., the matching procedure described in Tversky et al., 1988). However, experimental evidence suggests that the matching procedure tends to overestimate the indifference point (Bostic et al., 1990; Tversky et al., 1990), and that the choice-based procedure serves as a more reliable alternative. In the choice-based procedure, a participant is presented with pairs of prospects in each trial — one of which includes the target — and is asked to choose a preferred option. The target is then updated based on the participant's choice and presented in the next iteration until the procedure terminates. By analyzing the participant's choices, especially the level of the target where the preference switches, it is possible to identify the indifference point.

doi:10.6342/NTU202501601

The bisection method introduced by Abdellaoui (2000) is an example. It was first applied to eliciting the weighting function, and subsequently used in various aspects, mostly on the measurement of prospect theory (Abdellaoui et al., 2007; Abdellaoui et al., 2016; Bleichrodt and L'Haridon, 2023). In the bisection method, an interval is first identified that is expected to cover the indifference value. During each iteration, the midpoint of the interval is presented to the participant, and the interval is subsequently halved based on their choice. For example, if the option with current stimulus is preferred, the upper bound of the current interval is then lowered to current midpoint; otherwise, the lower bound is adjusted accordingly. After several iterations, the final working interval narrows down with a small width and one can estimate the indifference point with its midpoint.

However, little justification or theoretical basis about the bisection method has been made. In this thesis, we highlight two potential error sources the bisection method has overlooked. The first arises from the randomness inherent in participants' choice responses, and the second is related to the determination of the initial boundary of the interval. These problems might introduce errors, which would be propagate to subsequent stages in the experimental procedure.

To address to these issues, we introduce some psychophysical methods for threshold estimation as alternatives to elicit indifference points. These methods include ASA, PEST and MOBS. To inspect and compare the performance between these methods, two simulation studies are conducted with different settings. We choose the experiment in Abdellaoui et al. (2016) as the benchmark to perform simulations and to make comparisons, the reason being that their experiment comprehensively inspects prospect theory and loss aversion without parametric assumptions. Furthermore, the bisection method is mostly utilized in their other experiments with similar procedures.

doi:10.6342/NTU202501601

The experimental procedure of Abdellaoui et al. (2016) was also adopted in a study we had conducted to examine loss aversion, in which we utilized a simplified version of the bisection method for eliciting indifference points (see Chapter 3 for details). The existing (unpublished) data we had collected were analyzed and compared to two previous studies: Abdellaoui et al. (2016) and Bleichrodt and L'Haridon (2023). While our results share similarities with findings in the literature, we encountered a challenge with our simplified bisection — namely, that the obtained sequence of indifference points sometimes violate an expected non-decreasing order, which we refer to as monotonicity.[1] This condition should hold under the current experimental procedure, assuming that the underlying value function is strictly increasing. This further motivates the present study. By comparing different elicitation methods, it may provide extra insights into our existing data and possibly offer guidance for further research.

This thesis is structured as follows. First, in Chapter 2, we review the theoretic foundation of prospect theory and introduce the experiment of Abdellaoui et al. (2016) as a basis for our simulation settings. The analysis of existing data is also presented in Chapter 2. Chapter 3 introduces the bisection method with other adaptive psychophysical methods in the context of eliciting indifference points. The simulation settings across two studies are described in Chapter 4, and the results are presented in Chapters 5 and 6, respectively. Finally, in Chapter 7, we sum up the results, provide suggestion on the choice between methods, and discuss the limitations and further extensions about the present study.

[1] Indeed, our experimental results show a high proportion of participants violating monotonicity.

doi:10.6342/NTU202501601

doi:10.6342/NTU202501601

Chapter 2 Prospect Theory

2.1 Review of Prospect Theory

Kahneman and Tversky (1979) introduced the prospect theory as a framework to describe how individuals evaluate risky prospects. Unlike traditional expected utility theory, prospect theory posits that individuals' decisions are influenced by two key principles: (1) outcomes are evaluated based on gains or losses (relative to a reference point), rather than final assets; and (2) the value of each outcome is not weighted by probability, but by a decision weight. To further generalize the original prospect theory, Tversky and Kahneman (1992) later proposed the cumulative prospect theory (CPT) that can be applied to studying prospects with uncertainty (i.e., the probability an event may happen is unknown) and also prospects with any finite number of outcomes.

For instance, consider a binary risky prospect: $(x, p; y)$, which has probability p yielding a payoff x and a probability $1 - p$ yielding a payoff y. The outcome (x or y) is referred to as a gain if it is greater than the reference point x_0, and as a loss if it is lesser than x_0. A prospect is called mixed if it contains both gains and losses. A gain prospect contains no losses, while a loss prospect contains no gains. In the notation $(x, p; y)$ for a mixed prospect, $x > y$; for a gain prospect, $x \geq y$; and for a loss prospect, $x \leq y$.

If the $(x, p; y)$ is a mixed prospect, its CPT valuation is given by

$$V((x, p; y)) = w^{+}(p)v(x) + w^{-}(1 - p)v(y).$$

doi:10.6342/NTU202501601

If $(x, p; y)$ is a gain prospect or a loss prospect, its CPT valuation is given by

$$V((x, p; y)) = \begin{cases} w^+(p)v(x) + (1 - w^+(1-p))v(y), & x \geq y \geq x_0, \\ w^-(p)v(x) + (1 - w^-(1-p))v(y), & x_0 \geq y \geq x, \end{cases}$$

where $v(\cdot)$ is the value function with $v(x_0) = 0$; it represents the relationship between outcome and its psychological evaluation. And $w^+(\cdot), w^-(\cdot)$ are the weighting functions for gains and losses respectively; they represent the decision weight of actual probability of an event.

2.2 Measuring Prospect Theory

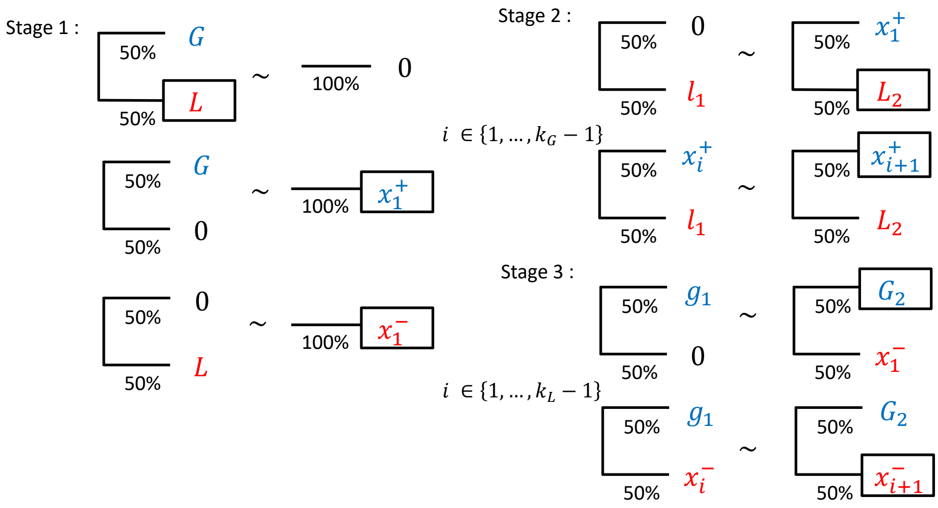
In applying CPT to real-world behavior, it is often essential to measure individual values and weighting functions (for a review of different approaches in measurement, see Fox and Poldrack, 2014). For instance, based on the trade-off method, Abdellaoui et al. (2016) proposed an experiment to measure the entire value function across gains and losses, resulting a standard sequence that is equally spaced on the value function. The method makes no assumptions about the parametric form of the value function and does not require measuring any decision weights. A key advantage of this method is that it enables the examination of loss aversion, which requires to elicit both gains and losses simultaneously. Additionally, the method can be extended to measure the value function in decisions under uncertainty, making it possible to observe CPT completely.

The procedure is sketched in Figure 2.1. In each trial, the participant is presented with a pair of prospects (or a pair containing a prospect and a sure value), and the goal is to obtain the target value (see the solid-lined rectangle in Figure 2.1) that makes both

doi:10.6342/NTU202501601

Figure 2.1

A Sketch of the Experimental Procedure in Abdellaoui et al. (2016)



Note. G, l_1, and g_1 are predetermined parameters and solid-lined rectangles represent target values to be estimated.

options be indifferent to him/her. The various iteration methods to approximate the target values—the main focus of the present study—will be explored in the next section.

The experiment has three stages, the first stage connects the domains of gains and losses, while the second and the third stages respectively elicit the standard sequence for each domain.

The first stage of the experiment has three parts, which would elicit L, x_1^+ and x_1^- respectively. For the first part, given a predetermined gain G, a loss L is elicited to make $(G, 0.5; L)$ indifferent to a certainty 0. Suppose 0 is the reference point ($v(0) = 0$), from the indifference relation $(G, 0.5; L) \sim 0$ we have:

$$w^+(0.5)v(G) + w^-(0.5)v(L) = v(0) = 0. \tag{2.1}$$

doi:10.6342/NTU202501601

A gain x_1^+ is then elicited to make $(G, 0.5; 0)$ indifferent to a certainty gain x_1^+. Next, a loss x_1^- is elicited such that $(L, 0.5; 0) \sim x_1^-$. Hence,

$$w^+(0.5)v(G) = v(x_1^+), \tag{2.2}$$

$$w^-(0.5)v(L) = v(x_1^-). \tag{2.3}$$

Combining (2.1), (2.2) and (2.3), one sees that the quantities of (L, x_1^+, x_1^-) elicited from the first stage satisfy

$$-v(x_1^+) = v(x_1^-). \tag{2.4}$$

The domains of gains and losses are then connected.

For the following stages, standard sequences were elicited in both gain and loss domains. In the beginning of the second stage, a predetermined loss $l_1 \in (L, 0)$ is given, and a loss L_2 would be elicited to make $(x_1^+, 0.5; L_2) \sim (0, 0.5; l_1)$. That is,

$$w^+(0.5)v(x_1^+) + w^-(0.5)v(L_2) = w^+(0.5)v(0) + w^-(0.5)v(l_1),$$

or

$$v(x_1^+) - v(0) = \frac{w^-(0.5)}{w^+(0.5)} \left(v(l_1) - v(L_2)\right). \tag{2.5}$$

Once we have L_2, then x_2^+ can be elicited with the indifference relation: $(x_2^+, 0.5; L_2) \sim (x_1^+, 0.5; l_1)$. Thus we have

$$v(x_2^+) - v(x_1^+) = \frac{w^-(0.5)}{w^+(0.5)} \left(v(l_1) - v(L_2)\right). \tag{2.6}$$

From (2.5) and (2.6), we have

$$v(x_2^+) - v(x_1^+) = v(x_1^+) - v(0). \tag{2.7}$$

doi:10.6342/NTU202501601

Equation (2.7) shows that the difference between the value of x_2^+ and x_1^+ equals to the difference between x_1^+ and 0. This implies that the values of x_2^+, x_1^+ and 0 are equally spaced. Moreover, by eliciting x_{i+1}^+ such that $(x_{i+1}^+, 0.5; L_2) \sim (x_i^+, 0.5; L_2)$, we can elicit the $x_3^+, x_4^+, ..., x_{k_G}^+$ with the following relationship:

$$v(x_{i+1}^+) - v(x_i^+) = v(x_1^+) - v(0), i = 1, 2, ..., k_G,$$

where k_G is the number of data points the experimenter aims to measure in the gain domain.

The third stage is similar to the second stage, but it intends to construct the standard sequence in the loss domain. Given a predetermined gain $g_1 \in (0, G)$, a gain G_2 is elicited to make $(G_2, 0.5; x_1^-) \sim (g_1, 0.5; 0)$. Once G_2 elicited, we can further determine $x_2^-, x_3^-, ..., x_{k_L}^-$ in a similar way to obtain

$$v(x_{i+1}^-) - v(x_i^-) = v(x_1^-) - v(0), i = 1, 2, ..., k_L,$$

where k_L is the number of data points the experimenter aims to measure in the loss domain. Very often, the experimenter would set $k_L = k_G = k$.

The above procedure elicits the standard sequence $\{x_k^-, ..., x_1^-, 0, x_1^+, ..., x_k^+\}$, which serves as a foundation for analyzing the value function. Additionally, the standard sequence enables the experimenter to examine loss aversion (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992), a key property of the value function in the CPT framework. Among several definitions of loss aversion, Abdellaoui et al. (2007) and Abdellaoui et al. (2016) highlighted that the definitions of Kahneman and Tversky (1979) and Köbberling and Wakker (2005) are empirically most useful. Based on these definitions, three indices can be computed to estimate loss aversion.

doi:10.6342/NTU202501601

Kahneman and Tversky (1979) defined loss aversion as $-v(-x) > v(x)$ for all $x > 0$. To operationalize their definition, $\frac{-v(-x_i^+)}{v(x_i^+)}$ and $\frac{-v(x_i^-)}{v(-x_i^-)}$ for $i = 1, \ldots, k$ is computed whenever possible. Usually, $v(-x_i^+), v(-x_i^-)$ are not directly observed, hence need to be determined by linear interpolation. Either its median or its mean can be used as an index to measure loss aversion. A participant is classified as loss averse if more than half of the values in $\left\{\frac{-v(-x_i^+)}{v(x_i^+)}, \frac{-v(x_i^-)}{v(-x_i^-)} \,\middle|\, i = 1, \ldots, k\right\}$ exceed 1, as gain-seeking if more than half are below 1, and as loss-neutral if none of these inequalities hold.

Köbberling and Wakker (2005) viewed loss aversion as a kink at the reference point. They defined loss aversion as $\frac{v'_\uparrow(x_0)}{v'_\downarrow(x_0)} > 1$, where $v'_\uparrow(x_0)$ is the left derivative of the value function at the reference point, and $v'_\downarrow(x_0)$ is the right derivative at the reference point. For the procedure in Abdellaoui et al. (2016), $x_0 = 0$, and x_1^+, x_1^- are the closest points to the reference point. This implies that $\frac{v'_\uparrow(x_0)}{v'_\downarrow(x_0)}$ can be approximated by $\left(\frac{v(0)-v(x_1^-)}{0-x_1^-} \middle/ \frac{v(x_1^+)-v(0)}{x_1^+-0}\right)$. Further, given that $v(x_1^+) = -v(x_1^-)$, the ratio then equals $\frac{x_1^+}{-x_1^-}$. A participant is classified as loss averse if the ratio $\frac{x_1^+}{-x_1^-}$ exceeds 1, as gain-seeking if it is below 1, and as loss-neutral if it equals 1.

2.3 Existing Data

The above-mentioned procedure has been used in Abdellaoui et al. (2016), Bleichrodt and L'Haridon (2023) and also in our previous (unpublished) study. This section summarizes the results from these studies. In particular, we focus on the loss aversion coefficients and the classifications, as these are the main analyses in these studies.

Our experiment parameters G, g_1, and l_1 were set as $+2000, +300, -300$, respectively, across experiments. However, there are two difference across these studies. First,

doi:10.6342/NTU202501601

the desired data points for each domain, k, is set as 6 in Abdellaoui et al. (2016) and 5 in Bleichrodt and L'Haridon (2023) and $k = 8$ in our study. To align with these studies while preserving as many data points as possible, the analysis below perform only using $\{x_6^-, \ldots, x_6^+\}$. Second, for the eliciting indifference in each data point, the first two studies utilized the "Bisection-Slider" method with four iterations, while we used the "SimpBisection" method with five iterations (The procedure of these elicitation methods are detailed in Chapter 3).

Additionally, in our earlier study, 87 participants were recruited, 14 of whom violated monotonicity in the standard sequence — specifically, cases where the sequence did not satisfy $x_6^- \leq \ldots \leq x_1^- \leq 0 \leq x_1^+ \leq \ldots \leq x_6^+$.[2] The proportion is higher than that reported in Abdellaoui et al. (2016) and in Bleichrodt and L'Haridon (2023) — 3 out of 75 participants and 10 out of 122 participants, respectively. For fear that these participants may provide noises to the data, we excluded them from the results below.

Table 2.1 summarizes the aggregate estimate of loss aversion coefficients and the classification of loss attitude from our data and previous research.[3] Across these studies, the majority of the participants were classified as loss-averse, and the median of loss aversion coefficient exceeded 1, regardless of the definition used. These results resemble the description of loss aversion in prospect theory. However, there are some differences between the results of these studies. For both definitions, our estimates are closer to the results of Bleichrodt and L'Haridon (2023): they exhibit smaller loss aversion coefficients and classify fewer participants as loss-averse compared to Abdellaoui et al. (2016) (Chi-squared test for homogeneity: $\chi^2(2, N = 257) = 9.1169, p = .01048$ from the def-

[2] Since the standard sequence satisfies $v(x_6^-) \leq \ldots \leq v(x_1^-) \leq 0 \leq v(x_1^+) \leq \ldots \leq v(x_6^+)$, the above inequality should be hold given that $v(\cdot)$ is strictly increasing.

[3] Abdellaoui et al. (2016) and Bleichrodt and L'Haridon (2023) did not specify whether mean or median values were used to compute the coefficient following the definition of Kahneman and Tversky (1979).

doi:10.6342/NTU202501601

inition of Kahneman and Tversky, 1979; $\chi^2(2, N = 266) = 7.8213, p = .02003$ from the definition of Köbberling and Wakker, 2005). The difference may be attributed to the heterogeneity of participants' loss attitudes across studies.

Table 2.1

Loss Aversion Coefficients and Classification Results

Definition	Study	Median [IQR]	Loss Averse	Gain Seeking	Loss Neutral
Kahneman and Tversky (1979)	Current (Mean)	1.33 [0.97 - 2.24]	49	22	2
	Current (Median)	1.32 [0.95 - 2.26]			
	Abdellaoui et al. (2016)	2.21 [1.06 - 5.52]	58	10	1
	Bleichrodt and L'Haridon (2023)	1.46 [0.82 - 2.96]	73	39	3
Köbberling and Wakker (2005)	Current	1.29 [0.91 - 2.00]	45	28	0
	Abdellaoui et al. (2016)	1.88 [1.06 - 4.50]	56	12	3
	Bleichrodt and L'Haridon (2023)	1.35 [0.95 - 2.55]	73	35	14

doi:10.6342/NTU202501601

Chapter 3 Methods for Eliciting Indifference Points

In the experimental procedure of Abdellaoui et al. (2016), it is crucial to elicit indifference points with high precision. This is because the experimental stimuli are sequentially chained — each elicitation task builds on responses from earlier ones. As a result, any error in the early estimates can propagate to subsequent trials, compounding over time and ultimately leading to biased measurements. To mitigate this issue, a choice-based procedure is employed for inferring indifference, as it tends to produce more consistent measurements compared to matching procedure (Bostic et al., 1990).

The following sections provide an overview of choice-based methods for eliciting indifference points. We first introduce the bisection method and some of its modifications, pointing out potential problems they may give rise to. We then explore methods commonly used in psychophysics and discuss their application to the domain of eliciting indifference points.

3.1 The Bisection Method

The bisection method has been used in research involving non-parametric methods for eliciting value and weighting functions (Abdellaoui, 2000; Abdellaoui et al., 2007). Its procedure is described below.

Suppose there is a pair of prospect, with one of them containing the target T that we want to elicit to make the pair indifferent. In the bisection method, one first needs to decide on an interval (a, b) that should cover T. At the beginning, the pair is presented

doi:10.6342/NTU202501601

with T as the midpoint of the interval $m_0 = \frac{a+b}{2}$, and the participant is asked to choose the preferred prospect. The interval is then updated according to the participant's choice. If the option containing T is chosen, meaning that option is more preferred, the updated interval containing the indifference point becomes (a, m_0). Conversely, if the other option is chosen, the interval is updated to (m_0, b). In the next trial, T is presented as the midpoint of the updated interval. This procedure repeats until a specified number of iterations is reached. The final presented T is then taken as the estimate of the indifference point.

The key part is how to determine the boundaries (a, b) at the beginning. Sometimes, it can be inferred by applying stochastic dominance to rule out cases where an option dominates the other. For example, when eliciting a certainty equivalent x_1^+ with the indifference relation $(G, 0.5; 0) \sim x_1^+$, where $G > 0$, stochastic dominance implies that if $x_1^+ \geq G$, it is preferred to $(G, 0.5; 0)$, and if $x_1^+ \leq 0$, it is not preferred. Hence, a reasonable boundary can be set as $(0, G)$. However, not all pairs can be used to determine both upper and lower boundaries through this condition. For example, when eliciting an outcome L with $(G, 0.5; L) \sim 0$, stochastic dominance only implies that L should be less than 0, otherwise one would strictly prefer the prospect option. In such cases, researchers have proposed two rules to determine the other boundary: one is to fix the interval width at a certain value (for example, 5000 in Abdellaoui, 2000), while the other is to fix the starting point of the target such that the pair have the same expected value (Abdellaoui et al., 2007; Abdellaoui et al., 2016). For the latter, the undetermined boundary can be specified in reverse, provided that the starting point is the midpoint of the boundaries.

While the bisection method efficiently narrows down the possible interval covering the indifference point, we propose two problems that may arise. First, it does not account for randomness in decision-making (Mosteller and Nogee, 1951). In the bisection proce-

doi:10.6342/NTU202501601

dure, the interval is narrowed with each update, and the discarded half is never revisited once a choice is made. This means that if a choice inconsistent with the underlying value function has made, the resulting estimate becomes biased, regardless of how many iterations will be conducted. Second, since the bisection method searching is done within a specified interval, it would be problematic if the interval is not selected properly at the beginning. If the interval does not cover the true value, the process cannot converge to the indifference point.

3.1.1 Bisection-Slider and SimpBisection

Two modifications of the original bisection method are presented and studied in this thesis: "Bisection-Slider" and "SimpBisection."

The Bisection-Slider method, which combines bisection and matching, was used in Abdellaoui et al. (2016) and Bleichrodt and L'Haridon (2023). In this procedure, the participant first faces $n-1$ iterations of bisections, then is asked to control a slider (scrollbar) to match two prospects. The possible range for response is three times wider as the interval resulting from the final bisection, which may address potential errors from previous choices. As the matching procedure may introduce bias and noises, the appropriateness of such combined method still needs to be validated. In a recent study, Bleichrodt and L'Haridon (2023) measured x_3^+ twice as a consistency check, from which it can be inspected that the difference between these measurements, the first to the third quartiles are $-500, 0$ and 10, respectively, and its standard deviation is 1150. It suggests that the Bisection-Slider method may provide noisy measurements.

The simplified bisection method (SimpBisection) is used in our previous research.

doi:10.6342/NTU202501601

Instead of maintaining the upper and lower boundaries in each iteration, this approach follows the inherent property of the bisection method as its update rule — the difference (step size) between the current stimulus and the next stimulus is halved in each iteration. In its implementation, an initial value is set to make the two options have equal expected values, and the initial step size is half the absolute value of this starting point. The next stimulus is determined by increasing or decreasing the current level by the step size, depending on the participant's choice, after which the step size is halved for each subsequent iteration.[4] Like the original bisection method, the procedure terminates after a specified number of iterations.

3.2 Adaptive Psychophysical Methods

To overcome the aforementioned problems that the bisection method may cause, we introduce some adaptive psychophysical methods as alternatives to elicit indifference points.

In psychophysics, a main topic is to studying the psychometric function, which describes the relationship between external stimuli and internal representation. For example, in a discrimination experiment, when a target stimulus x is paired with a fixed reference stimulus b, how likely is it that the participant will say x is greater than b? The relationship can be described by a psychometric function $Pr(x, b)$, the probability of choosing x over b. One aspect of psychophysics aims to find the stimulus strength that corresponds to a threshold in the psychometric function. For instance, one may seek to find x such

[4] This updating rule implicitly limits the possible range SimpBisection may explore. To see this, suppose the starting point is $S_0 > 0$, then the initial step size is $\frac{S_0}{2}$. Since the step size is halved at each iteration, the farthest points it may reach after N iterations are $S_0 \pm S_0 \sum_{k=1}^{N} (\frac{1}{2})^k$, which lie within $(0, 2S_0)$. A similar argument can be applied to the case where $S_0 < 0$.

doi:10.6342/NTU202501601

that $Pr(x, b) = 75\%$. This is called "threshold estimation." In this sense, eliciting indifference point is similar to a threshold estimation study with a 50% threshold, where the decision maker is equally likely to choose either option.

Traditional psychophysical methods for threshold estimation include the method of limits, the method of adjustment, and the method of constant stimuli. However, these methods have some deficits, one of which is a lack of experimental efficiency. Adaptive methods were developed to improve and speed up the experimental procedure (for a review, see Treutwein, 1995). Among these, we introduce three non-parametric methods: ASA, PEST and MOBS. Below, we briefly describe these methods.

3.2.1 ASA

"Accelerated Stochastic Approximation" (ASA, Kesten, 1958) is a modification of the Robbins-Monro process (Robbins and Monro, 1951) designed to enhance the speed of convergence. It has been proved that such a sequence converges to the threshold stimuli X_θ, provided that the underlying psychometric function is monotonic (Kesten, 1958).

Specifically, for a given stimulus level X_n at trial n, ASA updates the next stimulus level X_{n+1} according to the following rule:

$$X_{n+1} = \begin{cases} X_n - \frac{c}{n}(Z_n - \theta), & n = 1, 2, \\ X_n - \frac{c}{2+m}(Z_n - \theta), & n > 2, \end{cases}$$

doi:10.6342/NTU202501601

where Z_n is an indicator variable defined as

$$Z_n = \begin{cases} 1, & \text{if prospect with the target chosen,} \\ 0, & \text{otherwise.} \end{cases}$$

The parameter θ represents the target threshold and m denotes the number of choice reversals observed up to the current trial. Lastly, c is a constant, which can be thought as proportional to the initial step size. Note that, under the current setting where $\theta = 0.5$, the step size is symmetric in both directions, and the constant c is twice the initial step size.

The procedure stops when the step size is smaller than a specified minimum step, and the final estimate is the last tested level.

3.2.2 PEST

Taylor and Creelman (1967) introduced PEST (Parameter Estimation by Sequential Testing) to increase experimental efficiency in threshold estimations of the psychometric function. Subsequently, Bostic et al. (1990) applied PEST to determine the certainty equivalents of prospects, and Luce (2000) provided a summary of its rules in the context of estimating certainty equivalents. We adopt their modifications in our procedure.

In each trial, the target's outcome size decreases (or increases) by a step if the option was chosen (or not chosen) in the previous trial. Once an initial step size is established, the PEST procedure updates the subsequent step sizes according to the following rules:

1. Each time the choice is changed from the previous one, the step size is halved.

2. For the second consecutive step in the same direction, the step size remains un-

doi:10.6342/NTU202501601

changed from the previous trial.

3. Starting from the fourth consecutive step in the same direction, the step size doubles with each step, unless it exceeds the predefined maximum step size.

4. The size of the third consecutive step in a given direction depends on the sequence of steps preceding the most recent reversal: (1) If the step immediately before the reversal is a result of doubling, the third step is not doubled. (2) Conversely, if the step before the reversal is not doubled, the third step is doubled to the second step.

Similar to ASA, the PEST procedure stops when the step size attains predetermined minimum step size, and the final estimate is the last tested stimulus level.

3.2.3 MOBS

Inspired by the binary search in searching an ordered array, Tyrrell and Owens (1988) introduced MOBS (Modified Binary Search). While the method shares similarities with the bisection method, additional rules are incorporated to account for response errors. However, Treutwein (1995) criticized that these rules are heuristic and lacking of theoretical justification.

The rules of the MOBS procedure, adapted to the context of the current experiment, are outlined below. One can also refer to Anderson and Johnson (2006) for an example.

1. The range in which the target T is located is defined by two boundaries (a, b). Each boundary is implemented by a three-element stack: the high-stack and the low-stack. The top element of each stack represents the current boundary value, while the lower elements store previous boundaries. Initially, all elements of the high-

doi:10.6342/NTU202501601

stack are set to represent the upper bound of T, and the elements of the low-stack are set to lower bound of T. In each trial, one of these stacks is updated.

2. The target is presented as the midpoint between the top element values of the two stacks unless the alternative rule is applied.

3. With each response, one of the boundaries is updated. In line with the bisection method, if the option containing T is chosen, the current target value is added to the top of the low-stack. Conversely, if the option not containing T is chosen, the value is added to the top of the high-stack. For the updated stack, the previous elements are shifted down one position, and the bottom element is discarded.

4. When two consecutive choices are identical, the next stimulus is the top element from one of the stacks: if the option containing T is chosen twice consecutively, the next stimulus is the top element of the low-stack; if the option not containing T is chosen, the next stimulus is the top of the element of high-stack. If the new presentation is inconsistent with the participant's previous choices, the stack undergoes a process called *regression*.

5. Regression involves shifting all elements of a stack upward, discarding the top element, and replacing the bottom element with the initial boundary specification.

6. This process continues until the following two termination criteria are met: (1) A specified number of reversals in choices have occurred. (2) The last is smaller than 5% of the total measuring range.

As Tyrrell and Owens (1988) acknowledged, the 5% stopping criterion in the final rule is arbitrary. Alternatively, it can be aligned with the stopping rules of PEST or ASA by specifying a minimal step size.

doi:10.6342/NTU202501601

Chapter 4 Simulations

To study and compare the performance between different elicitation methods, we conducted a series of Monte Carlo simulation. In the simulation, decisions were modeled as an ideal agent following CPT with a choice function that incorporates the randomness of choice. The agent's underlying value function was predetermined and hence known. This simulated agent was then paired with various indifference-elicitation methods, repeating the experimental procedure described in Abdellaoui et al. (2016) for 1000 times. Also following their settings, the experiment parameters G, g_1, and l_1 were set as $+2000, +300, -300$, respectively; the number of data points for each domain was set as $k = 6$.

The sections below describe the simulation details about the ideal agent, including the underlying preference and choice mechanism, and the implementation of elicitation methods.

4.1 Ideal Agent

Stott (2006) has suggested a combination of power value function and logistic choice function be the best parametric form for modeling under CPT with stochastic choices. Hence we designed the ideal agent with such combination.

doi:10.6342/NTU202501601

First, the value function is a pairwise power function:

$$v(x) = \begin{cases} x^{\alpha}, & x \geq 0 \ , \\ -\lambda |x|^{\beta}, & x < 0 \ , \end{cases}$$

where the parameters α and β capture the curvature of the value function in the gain and loss side, respectively, and λ, the loss aversion coefficient, indicates the steepness of value function in the loss domain compared to the gain domain. For these parameter values, we followed the estimates from Tversky and Kahneman (1992) and set them to $\alpha = \beta = 0.88$ and $\lambda = 2.25$.

For the weighting function, since $p = 0.5$ is the only possible risk in the experiment, instead of assuming any functional form, we simply set $w^{+}(0.5) = w^{-}(0.5) = 0.5$.

Lastly, consider the choice function. Suppose an agent is evaluating two prospects, A and B, with respective values $V(A)$ and $V(B)$ determined by CPT. We modeled the agent's choice using a logistic function, such that the probability of choosing A is

$$\Pr(A) = \frac{1}{1 + e^{-\phi(V(A) - V(B))}} \ ,$$

where $\phi > 0$ is a parameter that captures the degree of randomness in the agent's decision-making process. A higher value of ϕ indicates that the agent's choices are more deterministic, closely following the value difference between the two prospects. Conversely, a lower value of ϕ implies greater randomness in the agent's choices, making them less sensitive to the value difference.

In previous research, the median of estimates of ϕ ranged from 0.06 to 0.25 (Nilsson et al., 2011; Glöckner and Pachur, 2012; Scheibehenne and Pachur, 2015). However, these

doi:10.6342/NTU202501601

studies employed prospects with smaller stakes than those used in the current experiment. Given the potential scale dependence of the power value function, these prior estimates of ϕ may not directly apply to the present paradigm. Instead of relying on these estimates, we set ϕ based on the first trial of our experiment. In this trial, the agent faces a choice between $(2000, 0.5; -2000)$ and 0. We specified the value of ϕ such that the probability of making an error (i.e., selecting a prospect with the lower CPT value) in that trial is 10^{-8}. This yields a baseline value of $\phi = 0.0367$. To capture individual differences in choice consistency, three levels for this parameter were used: $\phi = 0.367$, $\phi = 0.0367$, and $\phi = 0.00367$, corresponding to high, moderate, and low consistency, respectively.[5]

As a remark, in the bisection-slider method, the ideal agent occasionally needs to directly report its indifference value. To account for response error, we modeled the agent's response as following a normal distribution with a mean equal to the true indifference value and a standard deviation of 20.

4.2 Elicitation Methods

Table 4.1 summarizes the elicitation methods and their variants used in our simulation. Based on their underlying logic, these methods are generally classified into two categories: the bisection-based methods that maintain an interval intended to cover indifference points, and the non-bisection methods that do not. The bisection-based methods include Bisection, Bisection-Slider, and MOBS, while the non-bisection methods consist of ASA and PEST. However, SimpBisection does not fully align with either category, as it neither maintains an interval nor requires specification of an initial step size. Therefore,

[5]To quantify how these levels reflect choice consistency: when $\phi = 0.367$, the error rate in the first trial is below 10^{-15}, whereas for $\phi = 0.00367$, it exceeds 0.1.

doi:10.6342/NTU202501601

we treat SimpBisection as a distinct method, independent of the two main categories. Key features and parameters for these elicitation methods include: (1) the boundary-setting rule for the bisection-based methods or the initial point for the non-bisection methods and SimpBisection; (2) the initial step size for the non-bisection methods; and (3) the stopping criterion. These features and its realization in the simulation study are described below.

Boundary setting rule/Initial point: For the bisection-based methods, once the boundary is determined, the starting point is the midpoint of the interval. Two rules to determine the initial boundaries were applied in the simulation study: one fixed the boundary width to 5000 (Fixed Boundary Width), and the other set the pair of prospects with equal expected value (Equal-Expectation).

For the non-bisection methods and SimpBisection, instead of determining boundaries, the initial value needs to be specified at the beginning. This value can be assigned in a deterministic manner or with added fluctuation to avoid the influence of the starting position. Accordingly, two variants were proposed: one assigned the initial point such that the two prospects have the same expected value (Equal-Expectation), while the other introduced randomness by adding noise drawn from a normal distribution with a mean of zero and a standard deviation of 100, rounded to the nearest multiple of 5 (Random Initial Points).

Initial step size: For the bisection-based methods, this is determined by the initial interval range. For the non-bisection methods, this should be specified: it was set to 320 for PEST and ASA in current simulation. In SimpBisection, the initial step size was the absolute value of the starting point divided by 2.

Stopping criterion: We mentioned in the previous chapter that the stopping criteria

doi:10.6342/NTU202501601

between these methods do not agree. In the simulation, we considered two kinds of stopping criterion. In Study 1, we followed the convention used in the non-bisection methods: the process was stopped when its step size attains a minimum step, which was fixed at 5 (a very small value compared to the stakes). In Study 2, we adopted the bisection convention, in which the process stopped after a specified number of trials, which was set to either 5 or 10 separately.

Table 4.1

Summary of Elicitation Methods

Category	Method	Variant	Initial Step Size
Bisection-Based	Bisection Bisection-Slider MOBS	Equal-Expectation Fixed Boundary Width	*NA*
Non-Bisection	ASA PEST	Equal-Expectation Random Initial Points	*320*
SimpBisection		Equal-Expectation Random Initial Points	*abs(starting point)/2*

doi:10.6342/NTU202501601

doi:10.6342/NTU202501601

Chapter 5 Study 1

In Simulation Study 1, we aim to compare the performance of different elicitation methods. To ensure comparability under the same stopping criterion, we set all methods to terminate when the final step size is smaller than 5. Given the degrees of randomness ϕ at three levels — large ($\phi = 0.367$), medium ($\phi = 0.0367$) and small ($\phi = 0.00367$) — the agent repeated the experiment and the standard sequence were collected for each repetition.

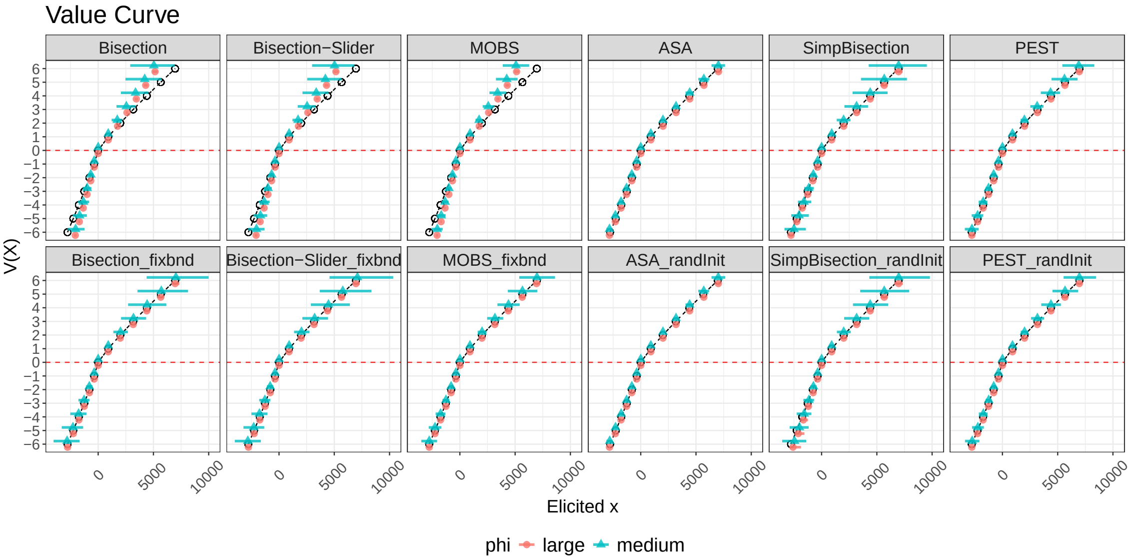
The aggregated results are presented in Figure 5.1. Each point represents the mean value from 1,000 repetitions for x_i^+ and x_i^-, $i = 1, ..., 6$. The extended line width indicates the range within which 95% of the simulation points fall, serving as a proxy for the 95% confidence interval (CI). The y-axis represents the corresponding values of $v(x_i^+)$ and $v(x_i^-)$. For simplicity, these values are assumed to be integers ranging from -6 to 6, as they are equally spaced. Each panel corresponds to a specific elicitation method. Different initialization variants for these methods are labeled with the suffixes *fixbnd* (indicating fixed boundary width for the bisection-based methods) and *randInit* (indicating random initial points for the non-bisection methods and SimpBisection). Methods without suffixes use equal-expectation scheme for boundary or starting point initialization. The black dashed line and hollow dots indicate the true values of the sequence $\{x_6^-, ..., x_6^+\}$. The plot omits the condition when ϕ is small for visualization purposes.

Two observations can be made from the plot. First, for the bisection-based methods, setting boundaries following equal-expectation scheme leads to biased estimates, whereas

doi:10.6342/NTU202501601

Figure 5.1

Overall Simulation Results for the Value Function



Note. The black dashed curves indicate the theoretical value function based on the specified parameters.

all other methods produce unbiased estimates. Second, even with initial boundary width fixed, the bisection-based methods still have wider 95% CI, compared to other methods.

Next, we examine the bias across different methods (and initialization schemes) for the three levels of ϕ. The results are shown in Figure 5.2. All methods are unbiased, except for the bisection-based methods using equal-expectation scheme at all levels of ϕ, and those using fixed boundary width scheme when the level of ϕ is small. The observed bias likely arises because, under equal-expectation scheme, the initial boundary is sometimes too narrow to capture the true parameter. However, this bias can be mitigated by setting a wider width, as in the fixed boundary width initialization scheme. Notably, when ϕ is small, the equal-expectation variant exhibits a smaller bias than the fixed boundary width variant. While this may initially seem counterintuitive, we propose that although the narrower boundary introduces bias, it also limits the scale of error introduced by "wrong"

doi:10.6342/NTU202501601

decisions. Hence, the equal-expectation variant produces consistent levels of bias across different settings of ϕ, and is less biased than the fixed boundary width variant when ϕ is small.

Another consideration is that one might suspect the bias is created by the specific specification of the value function. To address this issue, we varied the λ values in the underlying value function and ran additional simulations. The results, shown in Figure 5.4, demonstrate that the equal-expectation boundary setting still produces biased measurements when ϕ is large or medium.

In terms of statistical efficiency between different methods, Figure 5.3 shows the RMSE (Root Mean Square Error) for the different methods. When ϕ is large, all methods perform similarly well, having low RMSE values. However, when ϕ is medium or small, ASA achieves the lowest RMSE, followed by PEST, MOBS, Bisection and Bisection-Slider, and finally SimpBisection.

From these results, ASA appears to outperform all other methods, especially when the participants' choices involve higher levels of randomness. However, ASA has a significant drawback: it is time-consuming, requiring over 100 iterations per point to meet the stopping criterion (see Table 5.1). This is undesirable in empirical research. In contrast, Bisection, Bisection-Slider and SimpBisection require less than 10 trials per estimate, making them fastest and most efficient among all these other methods. Nevertheless, the current stopping rule may be unnecessarily strict for ASA. For example, with an initial step size of 320, it would take at least 64 iterations to meet the stopping criterion, even though ASA may have already converged earlier.[6] These findings motivate the

[6] We conducted additional analysis to explore this possibility. In this analysis, the procedure was set to terminate when 6 reversals occurred within the last 8 iterations. The results suggest that, on average, both variants of ASA require approximately 11, 17, and 19 iterations per point for large, medium, and small ϕ

doi:10.6342/NTU202501601

next simulation study, which examines the performance under conditions with a limited number of iterations.

Figure 5.2

Bias of Elicited x_i at Three Levels of ϕ in Study 1

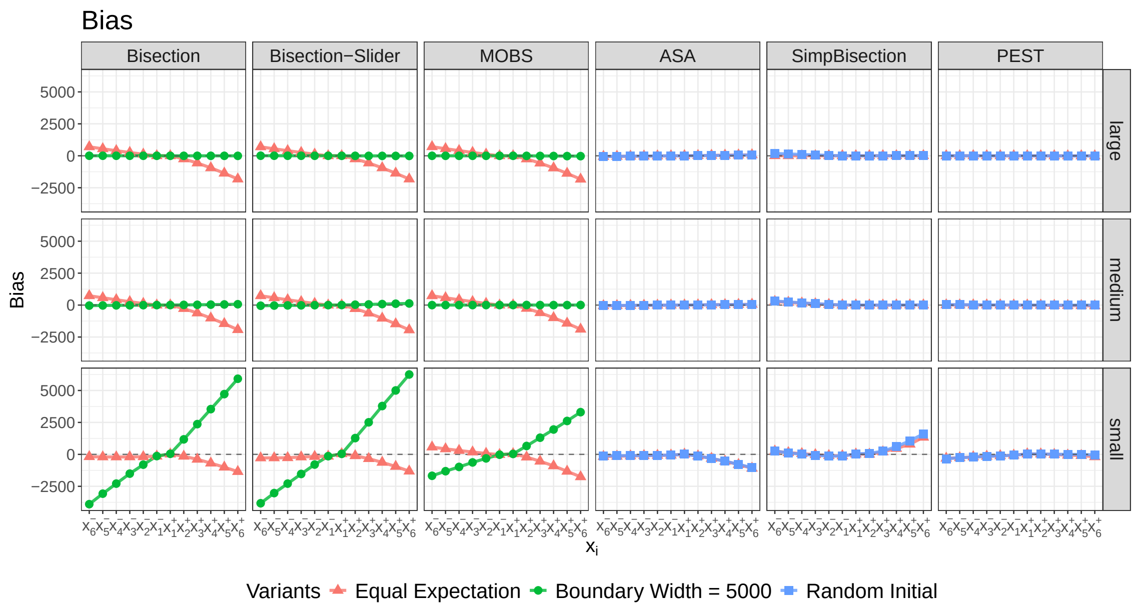
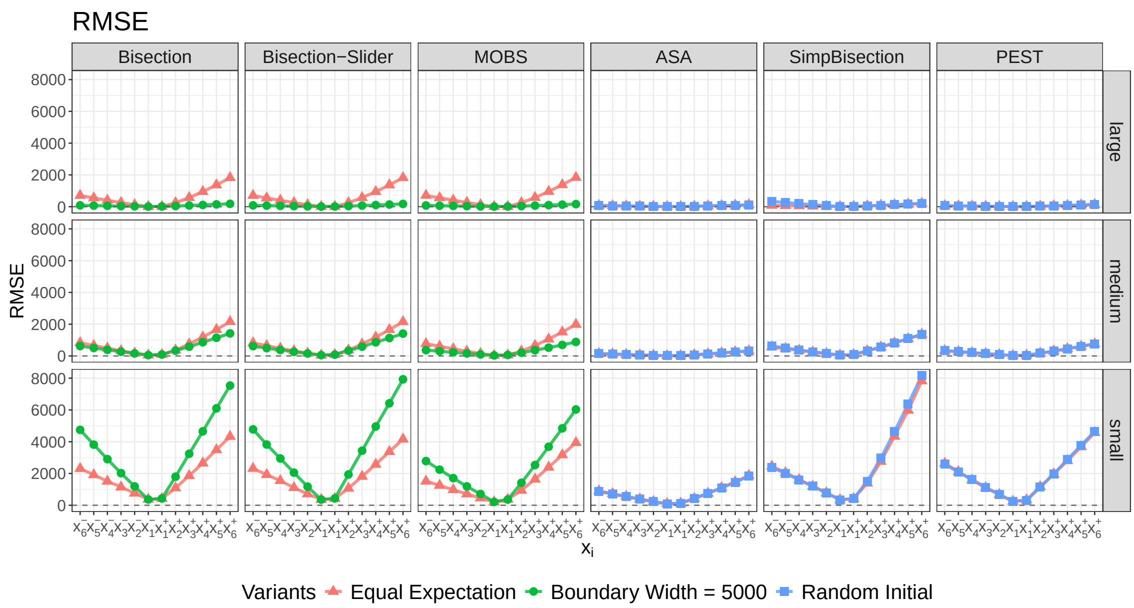


Figure 5.3

RMSE of Elicited x_i at Three Levels of ϕ in Study 1



---

levels, respectively. The resulting bias and RMSE were also similar to those obtained under the current criterion (see the ASA column in Figures 6.1 and 6.2).

doi:10.6342/NTU202501601

Table 5.1

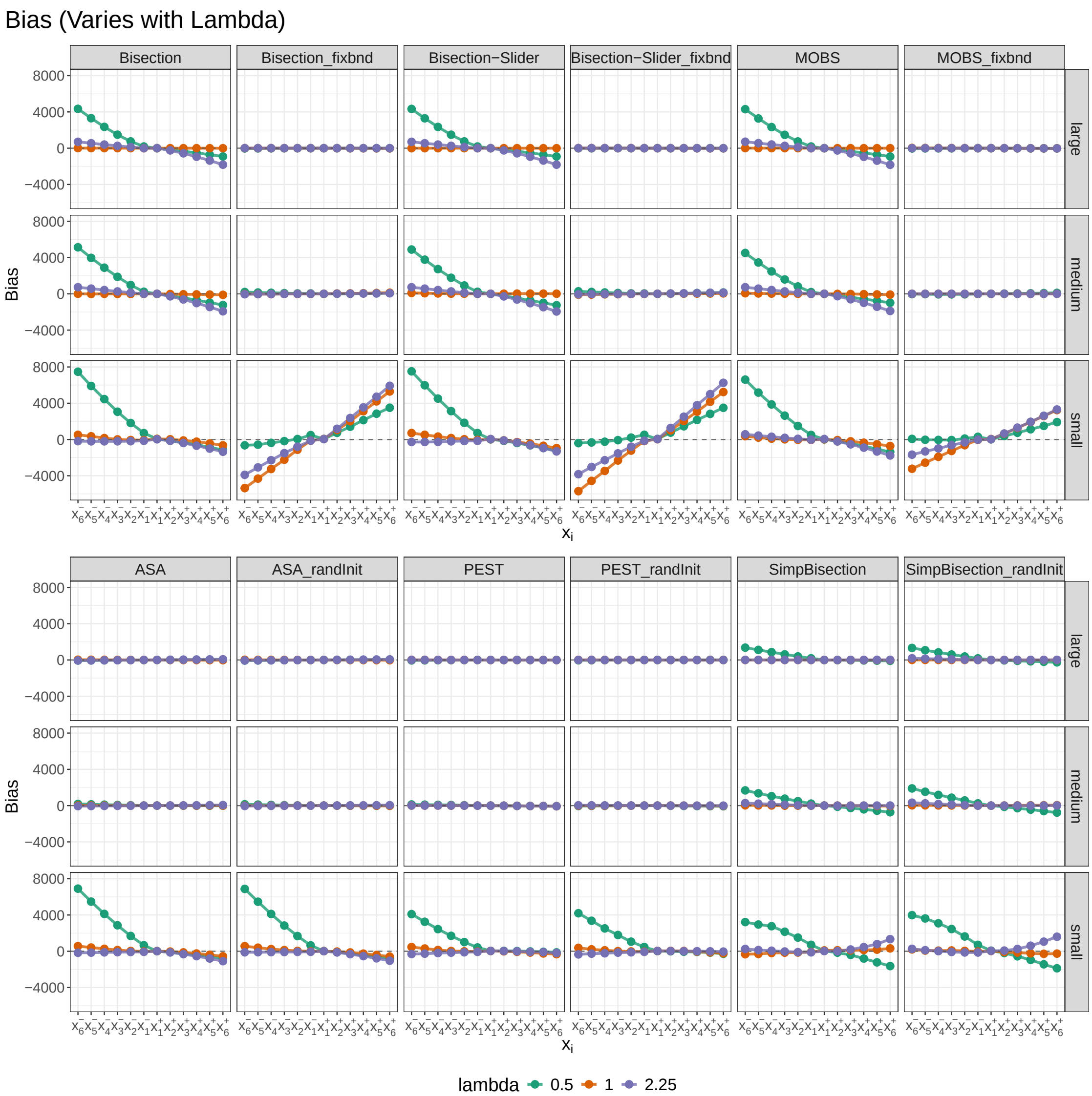
Mean Iteration Count for Each Point in Study 1

ϕ	Method	Mean Iteration Count
Large	Bisection	8
	Bisection_fixbnd	9
	Bisection-Slider	8
	Bisection-Slider_fixbnd	9
	MOBS	14
	MOBS_fixbnd	16
	ASA	117
	ASA_randInit	117
	SimpBisection	9
	SimpBisection_randInit	9
	PEST	16
	PEST_randInit	16
Medium	Bisection	8
	Bisection_fixbnd	9
	Bisection-Slider	8
	Bisection-Slider_fixbnd	9
	MOBS	16
	MOBS_fixbnd	21
	ASA	138
	ASA_randInit	138
	SimpBisection	9
	SimpBisection_randInit	9
	PEST	21
	PEST_randInit	21
Small	Bisection	7
	Bisection_fixbnd	9
	Bisection-Slider	7
	Bisection-Slider_fixbnd	9
	MOBS	16
	MOBS_fixbnd	23
	ASA	144
	ASA_randInit	144
	SimpBisection	9
	SimpBisection_randInit	9
	PEST	23
	PEST_randInit	23

doi:10.6342/NTU202501601

Figure 5.4

Bias of Elicited x_i, Varying in λ, in Study 1



doi:10.6342/NTU202501601

Chapter 6 Study 2

In Simulation Study 2, we set the stopping criterion to a limited iterations. The procedure terminated when a specified number of iterations (5 or 10) was reached, and the final presentation was used as the measurement. The bisection-based methods with equal-expectation initiation scheme were excluded due to their biasedness demonstrated in the previous simulation. To address the time-consuming issue of ASA, we also proposed a hybrid variant that combines Bisection with ASA ("Bisection-ASA" thereafter). During the early stage, the procedure began with the bisection method and switches to ASA when the step size became smaller than 500. The step size when ASA initiates was set to 100 (i.e., $c = 200$). Aside from these adjustments, the simulation settings were identical to those in Study 1.

Figure 6.1 shows the results of bias in Study 2.[7] In most cases, five iterations are insufficient to produce unbiased estimates for all x_i^+ and x_i^-, particularly for points farther from the reference point. This is possibly due to error propagation. Despite this, SimpBisection remains largely unbiased, even when only five iterations. When the iteration number is increased to 10, the methods generally become largely unbiased. Under large and medium ϕ, the maximum bias at a single point across all methods falls within ± 350. However, under small ϕ, the bisection-based methods and Bisection-ASA still yield biased estimates, even when 10 iterations are used. Lastly, for non-bisection methods and

[7] When the ϕ level is small, the simulated agent's choices become highly random, potentially leading to program errors in the MOBS and SimpBisection methods. As a result, these erroneous simulated data points were removed from the analysis. The most extreme case occurred with MOBS at five iterations, where 110 out of 1000 simulated results were discarded. In all other conditions, no more than 12 results were discarded, so this should not significantly impact the final analysis.

doi:10.6342/NTU202501601

SimpBisection, the initialization scheme appears to have little influence on bias.

Figure 6.2 illustrates the RMSE across different settings in Study 2. For visualization purpose, Figure 6.3 presents the relative RMSE compared to the bisection method. It is calculated as the ratio of RMSE of a method to that of the bisection method. A value less than 1 indicates that the method has a lower RMSE. It can be observed that when ϕ is large, Bisection and Bisection-Slider exhibit the lowest RMSE under both 5 and 10 iterations compared to all other methods. However, as the ϕ level changes to medium and 10 iterations are taken, Bisection-ASA and ASA has the smallest RMSE compared to other methods except for x_1^-, x_1^+, and ASA has the smallest RMSE when the ϕ level becomes small. Its relative RMSE are close to 1 when the ϕ levels are large and medium; under small level of ϕ, it is even lower than 1 at some points. Results also show that, although the procedures differ slightly, SimpBisection achieves similar — and in some cases even better — performance compared to Bisection.

As a note, the strong performance of Bisection-Slider under a small trial number might stem from the simulation setup, which produced a precise response during the matching task. The simulation result in Bisection-Slider may be spurious since past research has pointed out the bias in the matching procedure. In future study, empirical data are needed to validate the effectiveness of Bisection-Slider compared to other methods. In a recent article, Bleichrodt and L'Haridon (2023) conducted the same experiment with identical stakes as in the bisection-slider procedure. They measured x_3^+ twice as a consistency check. We computed the difference between repeated measures and estimate the standard deviation of the random error (the estimation details can be found in the Appendix). The estimate is 813.227, which is much larger than in our simulation setting.

doi:10.6342/NTU202501601

Figure 6.1

Bias of Elicited x_i at Three Levels of ϕ in Study 2

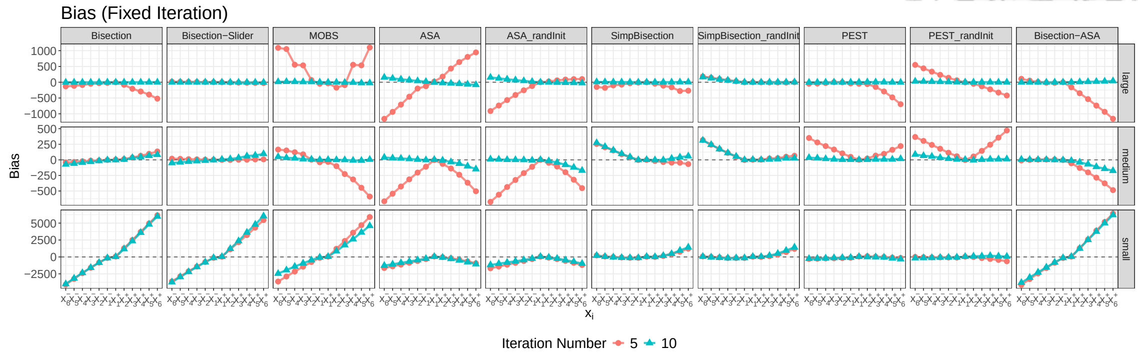


Figure 6.2

RMSE of Elicited x_i at Three Levels of ϕ in Study 2

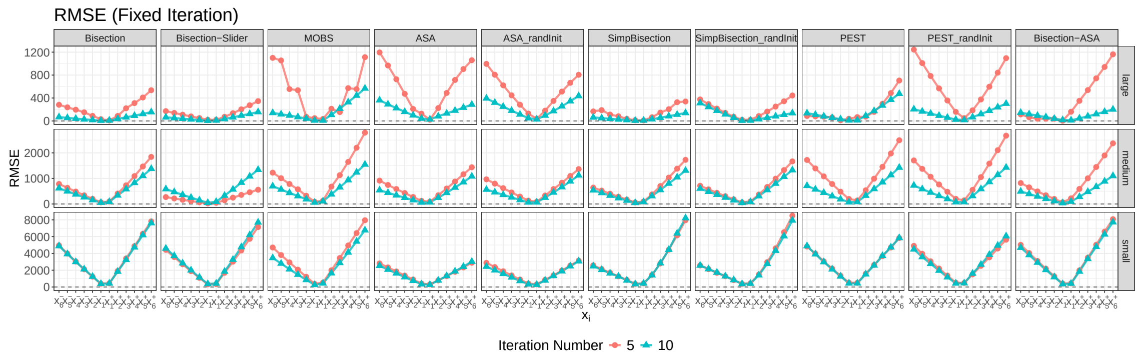
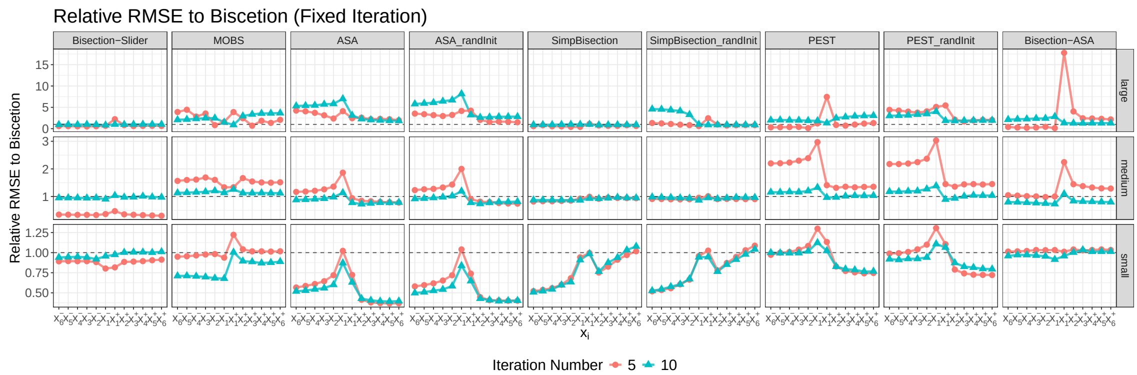


Figure 6.3

Relative RMSE of Elicited x_i at Three Levels of ϕ in Study 2



Besides the bias and RMSE of each points in the standard sequence, we also analyze three different loss aversion indices. Two of them follow the definition in Kahneman

doi:10.6342/NTU202501601

and Tversky (1979), computed from the mean or the median of $\frac{-v(-x_i^+)}{v(x_i^+)}$ and $\frac{-v(x_i^-)}{v(-x_i^-)}$ for $i = 1, ..., 6$. The other one follows the definition of Köbberling and Wakker (2005), that is, the ratio $\frac{x_1^+}{-x_1^-}$. These indices are named by their definition and computations, hence called "KT_mean," "KT_median" and "KW," respectively. Since the estimates of x_i^+ and x_i^- were generally biased across different methods with varying degrees of randomness ϕ when five iterations were used, it makes little sense to compute and compare the loss aversion estimates under such situations. In the text below, we only show the results where 10 iterations were used.

Figure 6.4 shows the mean of the estimates and the range where 95% of the simulation results fall, as a proxy of 95% CI. The black dashed line in the plot indicates the true loss aversion coefficient value according to that index. First, one can see when ϕ is large or medium, the estimates across different methods are unbiased under 10 iterations, whereas all methods yield more or less bias and a wider range of 95% CI when the ϕ level is small. For example, for PEST on the KW coefficient estimate, a great notion of 95% CI covers values less than 0, which is unreasonable. It can also be seen clearly in Figure 6.5.

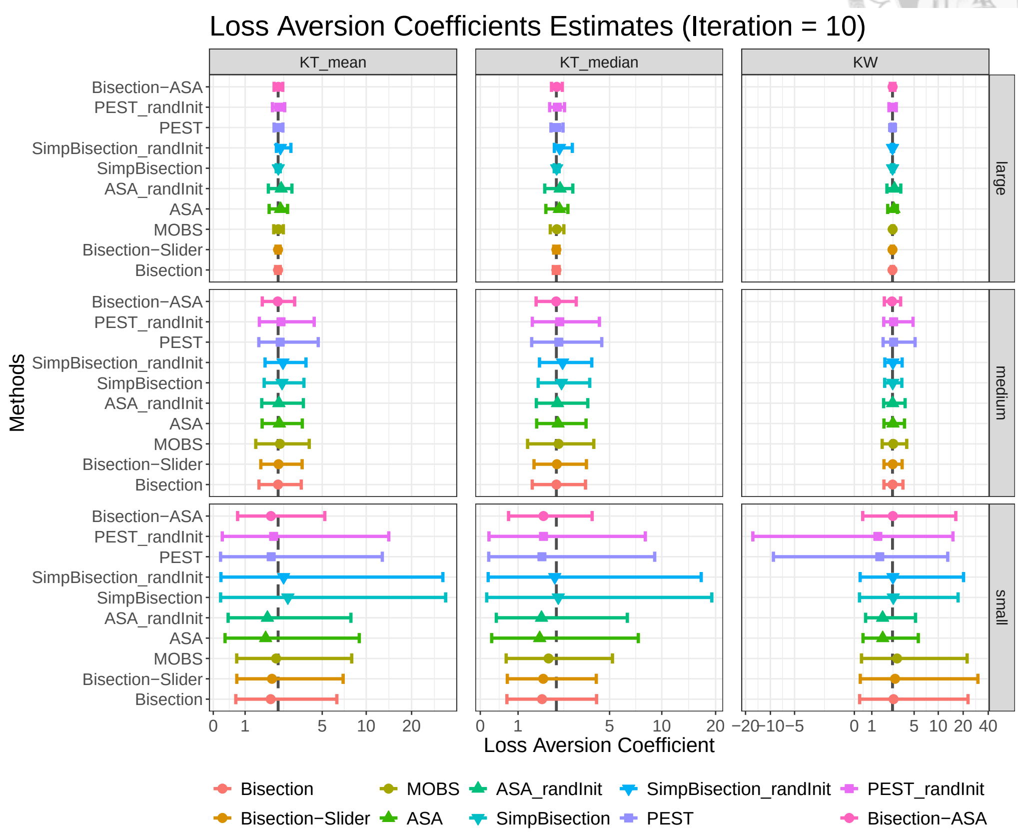
Second, in terms of efficiency, the RMSE results can be seen in Figure 6.6. While Bisection, Bisection-Slider and SimpBisection yield the lowest RMSE when ϕ is large, Bisection-ASA yields the lowest RMSE when ϕ is medium. When ϕ is small, the results are mixed. However, since all these methods are not very precise in this situation, it may mean less in discussing the performance between one to another. From the analysis of loss aversion coefficients, SimpBisection exhibits similar RMSE to those of the bisection method when ϕ is large or medium.

Lastly, we analyze the proportion of simulated data that violates monotonicity. Such

doi:10.6342/NTU202501601

Figure 6.4

Loss Aversion Indices Estimates and 95% Confidence Interval (Iteration = 10)



violations may raise concerns about whether participants fully understood the instructions or properly engaged with the experiment. Consequently, experimenters may exclude these cases from the final analysis. However, such violations could also stem from inherent choice randomness rather than true deviations in preference. Given the effort required to collect experimental data, minimizing these violations is crucial.

Table 6.1 summarizes the proportion that violating monotonicity across different methods excluding cases with no violations. For the bisection-based methods, the violation is very unlikely to occur, as stochastic dominance is assumed and the estimates are therefore constrained within a reasonable interval. Among the remaining methods, PEST and SimpBisection show higher violation rates, especially when ϕ is at small level. The combination of the randomness in participants' choice and the tendency of SimpBisection

doi:10.6342/NTU202501601

Figure 6.5

Bias of Loss Aversion Indices Estimates (Iteration = 10)

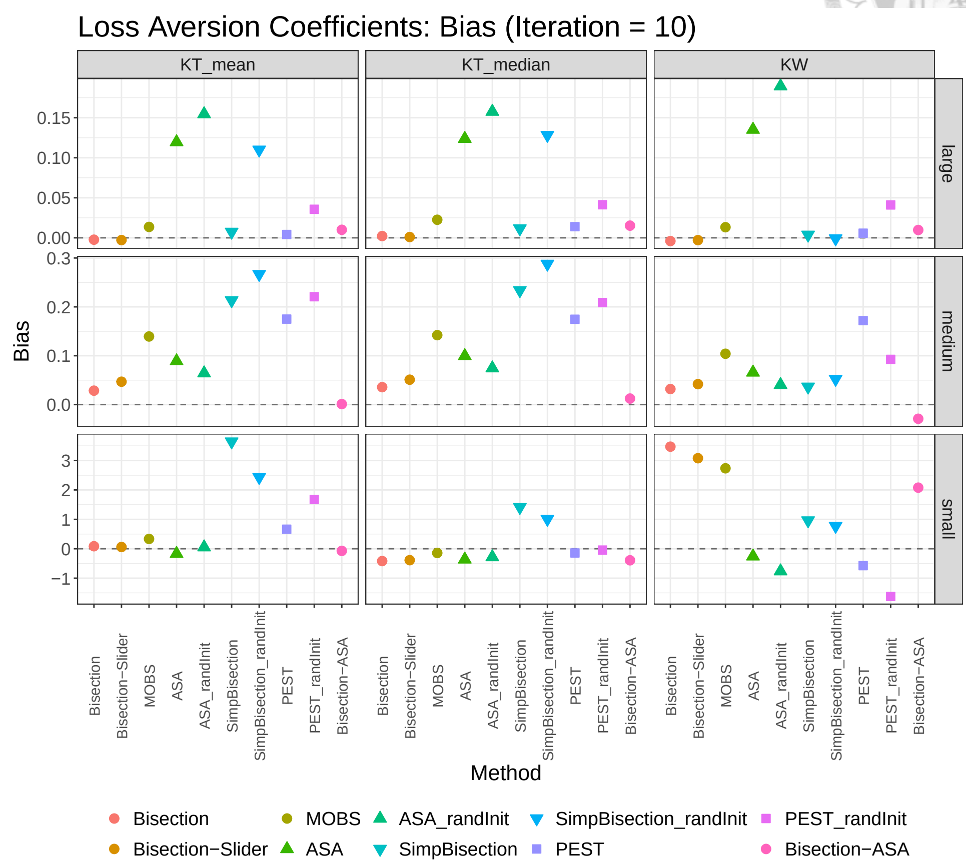
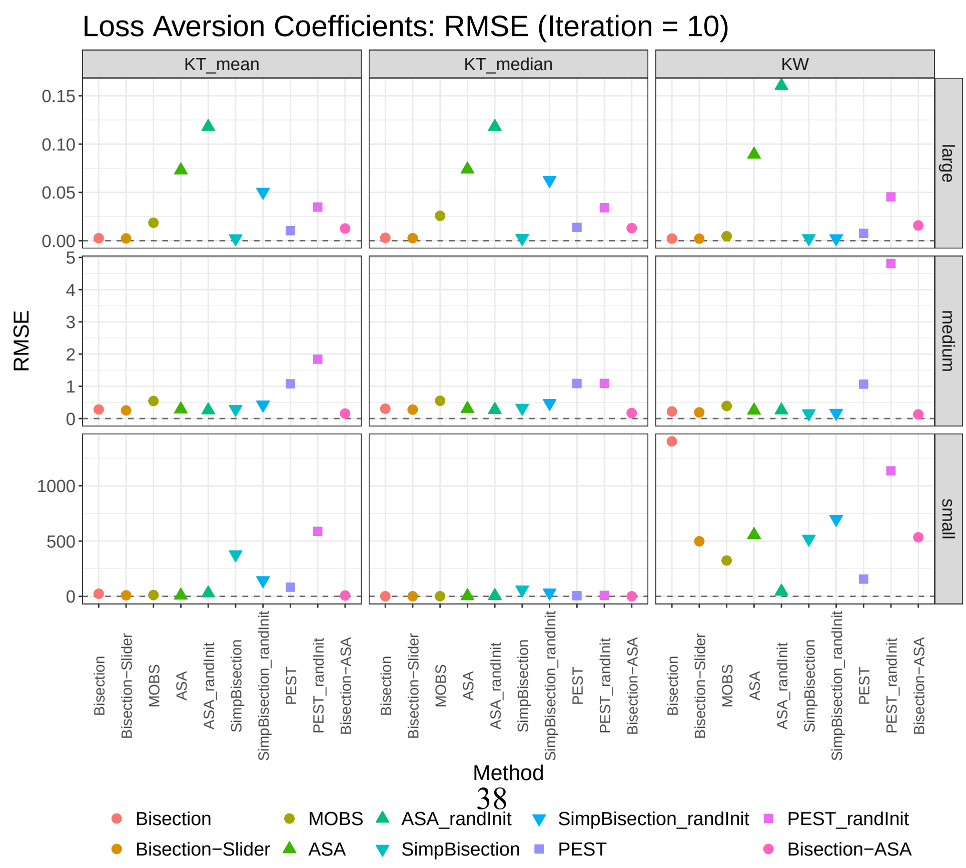


Figure 6.6

RMSE of Loss Aversion Indices Estimates (Iteration = 10)



to produce violations under fixed number of iterations may explain the increased violation rate in our (unpublished) data.

In conclusion, Study 2 shows that the bisection method performs well with a limited number of iterations when participants' choices are more deterministic. However, when choices involving higher randomness, ASA emerges as a better alternative. Under extreme conditions — where participants respond almost entirely at random — none of the methods yield precise estimates. The performance of the Bisection-ASA hybrid lies between the two methods: when Bisection performs poorly and ASA excels, or vice versa, Bisection-ASA generally achieves intermediate performance, making it a balanced compromise. SimpBisection performs similarly to the bisection method and can also serve as a viable alternative. These methods are recommended for the current experimental setting. That said, the final choice of method should still depend on the characteristics of the participants. The study also underscores the importance of participant screening during recruitment, as individuals with highly random response patterns are likely to produce noisy data regardless of the elicitation method used.

Lastly, regarding the number of iterations, using only five iterations is likely to yield biased estimates, even when participants are more deterministic (i.e., with a higher level of ϕ). Past research often fixed the number of iterations to determine the indifference point. For example, Abdellaoui (2000), Abdellaoui et al. (2007), and our study, used 5-7 iterations, while Abdellaoui et al. (2016), Bleichrodt and L'Haridon (2023) used only four iterations (three bisections plus one slider matching task). We recommend using more iterations per data point to obtain more precise estimates. However, we also understand that some experiments may need to limit a smaller number of iterations due to time or practical constraints. For researchers seeking a faster elicitation procedure, SimpBisection

doi:10.6342/NTU202501601

may be a good alternative, as it yields unbiased estimates across different levels of ϕ even with just five iterations.

Table 6.1

Proportion of Violations in Monotonicity

ϕ	Method	Iteration = 5	Iteration = 10
Large	PEST_randInit	6.2%	0%
Medium	ASA	0.1%	0%
	ASA_randInit	0.2%	0%
	SimpBisection	0.8%	0.1%
	SimpBisection_randInit	1.1%	0.3%
	PEST	14.6%	0.8%
	PEST_randInit	17.7%	1%
Small	ASA	31.8%	31.9%
	ASA_randInit	33.5%	33.1%
	Bisection-ASA	0%	1%
	SimpBisection	91.3%	90.2%
	SimpBisection_randInit	90.9%	86.6%
	PEST	70.7%	69.8%
	PEST_randInit	75%	69.1%

doi:10.6342/NTU202501601

Chapter 7 Discussion

In this thesis, we examined the bisection method introduced by Abdellaoui (2000) and some of its modifications, as well as several adaptive psychophysical methods, in the context of eliciting indifference points. Using the experimental setup of Abdellaoui et al. (2016) as a basis, we conducted two simulation studies under different settings to compare the performance of these methods.

Study 1 showed that when the number of iterations was not limited, all methods, except the bisection-based method using the equal-expectation boundary scheme, produced unbiased estimates. Among these methods, ASA had the lowest RMSE, albeit with the highest iteration count. Conversely, the bisection method achieved estimates with only one-tenth iterations to ASA, highlighting its experimental efficiency. Motivated by the efficiency of the bisection method, Study 2 constrained the number of iterations to 5 and 10. The results indicated that five iterations per estimate was insufficient for most methods to attain unbiased estimates. When the number of iterations was limited to 10, the bisection method performed best when participants' choices were relatively deterministic. ASA outperformed the bisection method in scenarios involving higher levels of randomness in participants' responses. Simulations also showed that our SimpBisection method performs comparably to the bisection method, and remains unbiased even when only five iterations are used.

The simulation results provide evidence supporting the use of the bisection method under some circumstances. With careful attention to boundary settings, the bisection

doi:10.6342/NTU202501601

method yields precise estimates within a limited number of iterations when participants' choice are not extremely random. For experiments involving participants with higher levels of choice randomness, ASA with more iterations emerges as a preferable alternative. A hybrid of the bisection method and ASA also performs well across different situations, it may serve as a conservative choice when the experimenter has little prior knowledge about participants' characteristics. Lastly, SimpBisection is recommended in experiments with only small number of iterations allowed.

For future works, empirical studies are crucial and necessary for further validation, as simulations cannot fully account for real-world situations. Such studies would not only test the robustness of these methods, but also help quantify the level of choice randomness in participants — capturing not just general variability but also time-varying patterns of inconsistency, such as those caused by fatigue or adaptation during the experiment. Empirical evidence could offer additional insights and provide better guidance for experimenters in selecting the method that best fits their needs.

Another possible future direction is improving the current methods. The performance of these elicitation methods may benefit from incorporating behavioral measures beyond mere choice responses. For example, response time may provide additional information about one's preference, particularly in stochastic choice behavior (Alós-Ferrer et al., 2021). Incorporating response time as an input may help algorithms to elicit indifference points more efficiently. Recently, Yang and Hsu (2024) generalized the ASA algorithm to incorporate response time, and demonstrated a lower MSE compared to the original ASA. Such modifications are worth investigating in future research.

Additionally, the boundary setting schemes for the bisection-based methods could

doi:10.6342/NTU202501601

be further refined. Rather than relying on fixed rules (as the two schemes we have investigated in the current study), adaptive boundary setting strategies may be explored. Experimenters could first explore the unbounded direction to identify the point where the participant reverses their response. This reversal point could then serve as a customized initial boundary. Currently, we are attempting to implement an adaptive boundary scheme in bisection-based methods. The procedure begins by setting the starting point using the equal-expectation approach. The algorithm then adjusts the target value by increasing or decreasing it in fixed step sizes (e.g., 500 units) until the agent reverses their choice, thereby revealing a preference change. The point at which this reversal occurs is used as a personalized boundary for the bisection-based elicitation methods. Preliminary results suggest that this adaptive scheme can reduce the large bias observed in bisection-based methods and Bisection-ASA, specifically under small ϕ condition within 10 iterations. It also achieves a lower RMSE than ASA under the same conditions. These results suggest the potential of the adaptive boundary scheme to enhance elicitation accuracy. However, this strategy requires several iterations for boundary identification, which may in turn increase the number of trials for elicitation. Despite this trade-off, the promising initial findings indicate that adaptive boundary setting is a worthwhile direction for further investigation to assess its robustness and practical viability.

Lastly, the methods examined in this thesis may also be extended to domains beyond eliciting the value function in prospect theory. For example, the bisection method has been employed in the non-parametric elicitation of the weighting function (Abdellaoui, 2000; van de Kuilen and Wakker, 2011). These experiments aim to elicit a probability that makes a participant indifferent between two prospects, rather than an outcome. Similarly, a well-known procedure developed by Holt and Laury (2002) for measuring risk

doi:10.6342/NTU202501601

aversion involves finding a switching probability between a risky, high-payoff prospect and a safe, low-payoff prospect. While the multiple price list technique is commonly used here, it has been criticized for its susceptibility to framing effects, central tendency biases, and the occurrence of multiple switching points—often considered indicators of low data quality (Andersen et al., 2006, see also Csermely and Rabas, 2016; Yu et al., 2021). Modified versions of ASA or bisection-based approaches could be adopted to address these limitations.

doi:10.6342/NTU202501601

References

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science*, *46*(11), 1497–1512. https://doi.org/10.1287/mnsc.46.11.1497.12080

Abdellaoui, M., Bleichrodt, H., L'Haridon, O., & van Dolder, D. (2016). Measuring loss aversion under ambiguity: A method to make prospect theory completely observable. *Journal of Risk and Uncertainty*, *52*(1), 1–20. https://doi.org/10.1007/s11166-016-9234-y

Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science*, *53*(10), 1659–1674. https://doi.org/10.1287/mnsc.1070.0711

Alós-Ferrer, C., Fehr, E., & Netzer, N. (2021). Time will tell: Recovering preferences when choices are noisy. *Journal of Political Economy*, *129*(6), 1828–1877. https://doi.org/10.1086/713732

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, *9*(4), 383–405. https://doi.org/10.1007/s10683-006-7055-6

Anderson, A. J., & Johnson, C. A. (2006). Comparison of the ASA, MOBS, and ZEST threshold methods. *Vision Research*, *46*(15), 2403–2411. https://doi.org/10.1016/j.visres.2006.01.018

Bleichrodt, H., & L'Haridon, O. (2023). Prospect theory's loss aversion is robust to stake size. *Judgment and Decision Making*, *18*, e14. https://doi.org/10.1017/jdm.2023.2

Bostic, R., Herrnstein, R., & Luce, R. D. (1990). The effect on the preference-reversal phenomenon of using choice indifferences. *Journal of Economic Behavior & Organization*, *13*(2), 193–212. https://doi.org/10.1016/0167-2681(90)90086-S

Csermely, T., & Rabas, A. (2016). How to reveal people's preferences: Comparing time consistency and predictive power of multiple price list risk elicitation methods.

doi:10.6342/NTU202501601

Journal of Risk and Uncertainty, *53*(2), 107–136. https://doi.org/10.1007/s11166-016-9247-6

Fox, C. R., & Poldrack, R. A. (2014). Prospect theory and the brain. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics* (2nd ed., pp. 533–567). Academic Press. https://doi.org/10.1016/B978-0-12-416008-8.00042-5

Glöckner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, *123*(1), 21–32. https://doi.org/10.1016/j.cognition.2011.12.002

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5).

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2).

Kesten, H. (1958). Accelerated stochastic approximation. *The Annals of Mathematical Statistics*, *29*(1), 41–59.

Köbberling, V., & Wakker, P. P. (2005). An index of loss aversion. *Journal of Economic Theory*, *122*(1), 119–131. https://doi.org/10.1016/j.jet.2004.03.009

Luce, R. D. (2000). *Utility of gains and losses: Measurement- theoretical, and experimental approaches*. Lawrence Erlbaum Associates Publishers.

Mosteller, F., & Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy*, *59*(5), 371–404.

Nilsson, H., Rieskamp, J., & Wagenmakers, E.-J. (2011). Hierarchical bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, *55*(1), 84–93. https://doi.org/10.1016/j.jmp.2010.08.006

Robbins, H., & Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, *22*(3), 400–407. https://doi.org/10.1214/aoms/1177729586

Scheibehenne, B., & Pachur, T. (2015). Using bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychonomic Bulletin & Review*, *22*(2), 391–407. https://doi.org/10.3758/s13423-014-0684-4

Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty*, *32*(2), 101–130. https://doi.org/10.1007/s11166-006-8289-6

doi:10.6342/NTU202501601

Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *The Journal of the Acoustical Society of America*, *41*(4A), 782–787. https://doi.org/10.1121/1.1910407

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, *35*(17), 2503–2522.

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

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, *95*(3), 371–384. https://doi.org/10.1037/0033-295X.95.3.371

Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *The American Economic Review*, *80*(1), 204–217. https://doi.org/10.1017/CBO9780511618031.009

Tyrrell, R. A., & Owens, D. A. (1988). A rapid technique to assess the resting states of the eyes and other threshold phenomena: The modified binary search (MOBS). *Behavior Research Methods, Instruments, & Computers*, *20*(2), 137–141. https://doi.org/10.3758/BF03203817

van de Kuilen, G., & Wakker, P. P. (2011). The midweight method to measure attitudes toward risk and ambiguity. *Management Science*, *57*(3), 582–598. https://doi.org/10.1287/mnsc.1100.1282

Wakker, P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge University Press.

Wakker, P., & Deneffe, D. (1996). Eliciting von Neumann–Morgenstern utilities when probabilities are distorted or unknown. *Management Science*, *42*(8), 1131–1150.

Yang, H.-H., & Hsu, Y.-F. (2024). The generalized Robbins–Monro process and its application to psychophysical experiments for threshold estimation. *Journal of Mathematical Psychology*, *120–121*, 102855. https://doi.org/10.1016/j.jmp.2024.102855

Yu, C. W., Zhang, Y. J., & Zuo, S. X. (2021). Multiple switching and data quality in the multiple price list. *The Review of Economics and Statistics*, *103*(1), 136–150. https://doi.org/10.1162/rest_a_00895

doi:10.6342/NTU202501601

doi:10.6342/NTU202501601

Appendix A — Supplemental Results in Study 2

A.1 Analysis of Bleichrodt and L'Haridon (2023) consistency check data

The data from Bleichrodt and L'Haridon (2023) repeatedly measured x_3^+ twice to serve as a consistency check. We utilized these data to estimate the variance of the random error involved in the Bisection-Slider task. Among the four variants of the experiment in the research, we only inspected the high-stakes experiment for risk to analysis since it matches our design in both stakes and the decision under risk setting.

To specify the error term, we assume that the value a participant would report in the slider task follows

$$S_{ij} = \tau_i + \epsilon_{ij} \ ,$$

where S_{ij} denotes the participant i's report in the j-th measurement in the slider task, τ_i is the true indifferent value of the participant i that matches two prospects, and ϵ_{ij} is the random error. For simplicity, we assume that all participants share the same degree of error, that is, $\epsilon_{ij} \sim N(0, s^2)$ and independently for each i, j.

The difference (Δ_i) for the same participant between both measurements of x_3^+ can be formulated as

$$\Delta_i = S_{i1} - S_{i2} = \epsilon_{i1} - \epsilon_{i2} \ ,$$

which has a variance of $2s^2$. Hence we can estimate s^2 by the sample variance of the difference between two measurements divided by 2.

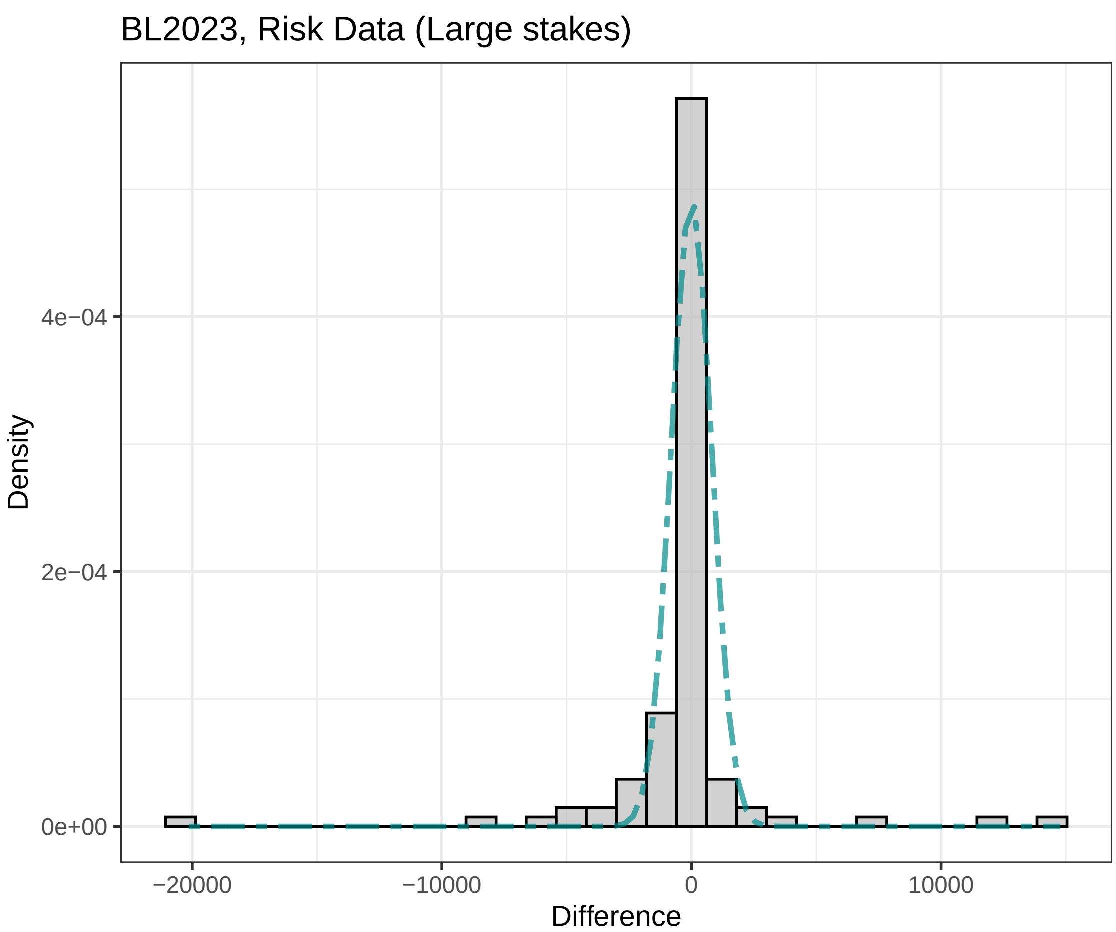
doi:10.6342/NTU202501601

However, in order to reduce possible noises, we did not include all participants' data in the estimation. First, we excluded the data which violates monotonicity. Second, the upper and lower 2.5% of the extreme value in the difference data were excluded. The remaining 106 data points were used to compute the estimate of s, leading to the estimate provided in the main text: $\hat{s} = 813.227$.

To help visualization, we provide the distribution of the difference data and a normal distribution curve of $N(0, 2\hat{s}^2)$ in Figure A.1

Figure A.1

Distribution of Difference Data in Bleichrodt and L'Haridon (2023)



doi:10.6342/NTU202501601