

東海大學經濟學系碩士班
碩士論文

決策個體在不確定下的選擇行為：
資訊反饋與學習

指導教授：戴中擎 博士

研究生：葉明璋 撰

中華民國九十九年七月

謝 辭

首先要感謝我的論文指導教授戴中擎老師。在寫作論文的過程中，恩師不斷的給予指導教誨，循序誘導，尤其老師總是能夠有耐心的解答我的疑惑，使我廣大的知識領域中，仍得以依循正確的方向，在此對於恩師對學生的投入與指正，致上最高謝意。和恩師的關係亦師亦友，也提供學生參與學術會議的機會，在投考博班時也給予學生許多的建議，不僅在學業上，對於人生的啓發更是受惠良多，感激之情，永難忘懷。

除此之外，由衷的感謝楊力行老師與池秉聰老師，撥冗在口試時對於學生觀念的澄清、論文寫作技巧與論文架構的指正，更提供了許多珍貴的建議，使我受益匪淺，更使得本論文臻至完善，內容更加的充實，在此對兩位老師表達我衷心的感激。

在過去所上的七百多個求學的日子，所上師長在學業上的細心教導，同窗摯友的情誼與鼓勵，尤其是與敬豪、秀娟、靜芳、婉如、金慶、士鴻、書庭、文鍵、依庭更是獲得了許多啓發。在經研所的生涯中，結識了許多來自各地的同學，很容幸在這段充實的時光有你們的陪伴扶持，讓我不覺得孤單，這將是我珍貴的回憶！也感謝冠華學長和游泳隊的朋友們在過去這段時間的關心。

最後感謝我的家人們，謝謝你們一直以來給予的支持與這二十多年來的養育之恩，讓我在求學間可以無後顧之憂，專心致至於學業上，順利完成碩士學爲位。

謹誌

於東海大學經濟研究所

民國九十九年七月

摘要

本文欲了解個體面對不確定的選項時選擇的行為，本文使用了 two-armed bandit 的問題來研究，我們參考了 J.J. Goodnow (1955) 的實驗比較了不同資訊回饋對受測者選擇行為的影響。在實驗裡受測者需要在畫面裡選擇“A”按鈕或是“B”按鈕，在做完選擇後受測者將會進入結果畫面。我們所設計的四個實驗分別在結果畫面有不同的資訊回饋或有不同的報酬計劃。利用受測者在做出轉換行為前得到“0”報酬的次數，利用這個方式建構一個截斷點策略 (cut point) 當作是受測者行為指標。再者又考慮到受測者決策時可能不管我們所給的資訊回饋，個體決策的差別可能只是轉換次數的多寡，因此以四個實驗裡所有受測者轉換次數的中位數當作分組的標準。

研究的結果顯示：幾個不同資訊回饋的平均報酬，在我們的實驗裡看不出差異，顯示更多的資訊無法幫助受測者選擇有較高報酬的選項導致使平均報酬沒有差異。在轉換次數方面，轉換次數在受測者之間的差異很大但只有實驗二、三和實驗三、四的平均轉換次數不同。表示資訊回饋不同對受測者不確定行為上的轉換行為上影響不大。在實驗裡轉換次數的趨勢是會隨著嘗試 block 數增加而減少。另外我們檢視了兩個實驗的截斷點策略，四個實驗裡的受測者在這個表現上沒有一致的趨勢。最後我們把受測者分成了高轉換組與低轉換組之後，截斷點的趨勢在 block3 之前是相同的。資訊回饋對個體在不確定行為下的選擇影響沒有差異。個體行為受到個體本身差異的影響較多，相對的不同資訊回饋影響較小。

關鍵字: two-armed bandit、資訊回饋、開發、探索、實驗經濟學、不確定下選擇行為

ABSTRACT

The study wants to comprehend individual choosing behavior under uncertainty, by using two-armed bandit problem. We compared different feedbacks impact on subjects choosing behavior. In four experiments subjects decide click button “A” or button “B”, then subjects will enter the result phase. Four experiments have different feedbacks in result phase or different payoff schedules. We use times of button “A” being chosen payoff · switch number and cut point as indexes measuring behavior. The cut point index is constructed by counting how many times payoff “0” occurred between each switches. Also, considering that individual may ignore feedbacks that we gave. People using different strategies may simply a result of they have different switching frequency. We addressed this by using the median that switching times in all four experiments to category our subjects.

Results showed that different feedbacks have no effect on individual choosing behavior under uncertainty. Instead, we found that difference among individuals has effect on choosing behavior under uncertainty.

Keywords: two-armed bandit · information feedback · exploration · exploitation · experimental economics · choice under uncertainty

目錄

一、緒論.....	1
1.1 研究背景與動機.....	1
1.2 研究問題.....	2
1.3 研究目的.....	3
二、文獻回顧.....	4
2.1 Two-armed bandit 實驗.....	4
2.2 Two-armed bandit 實驗結果與模型的對照.....	9
2.3 Two-armed bandit 資訊的獲取與學習.....	15
三、實驗設計.....	18
3.1 實驗概述.....	18
3.2 實驗參數.....	19
3.2.1 實驗一.....	21
3.2.2 實驗二.....	22
3.2.3 實驗三.....	23
3.2.4 實驗四.....	23
3.3 使用軟體.....	24
3.4 實驗流程.....	24
四、實驗結果.....	27
4.1 平均表現.....	27
4.2 選擇的狀況.....	28
4.3 轉換次數.....	29
4.5 轉換次數分組.....	34
4.5.1 平均表現.....	34
4.5.2 選擇的狀況.....	35
4.5.3 截斷點.....	37
五、結論與展望.....	39
參考文獻.....	41
附錄一 實驗使用之指導語.....	44
附錄二 截斷點結果.....	45
附錄三 簡單回歸模型結果.....	48
附錄四 轉換次數、選擇 A 次數與報酬的 Box-plot 圖.....	50

圖表索引

圖目錄

圖 1：實驗中選擇畫面.....	25
圖 2：實驗一 結果畫面.....	26
圖 3 實驗一到實驗四的平均報酬.....	27
圖 4 實驗一到實驗四選擇 A 的比例.....	28
圖 5 轉換次數.....	30
圖 6：四個實驗截斷點.....	31
圖 7 四個實驗的 atob -btoa 兩個數列相減的趨勢.....	32
圖 8 轉換數分組的報酬.....	34
圖 9 高轉換組與低轉換組選擇 A 的平均次數.....	35
圖 10 高轉換組與低轉換組對從 A 選項轉換到 B 選項的截斷點行爲	38
圖 11 實驗一至四 a to b 截斷點.....	45
圖 12 實驗一至四 b to a 截斷點.....	45
圖 13 實驗一截斷點.....	46
圖 14 實驗二截斷點.....	46
圖 15 實驗三截斷點.....	46
圖 16 實驗四截斷點.....	46
圖 17 轉換數分組後的 a to b 數列.....	47
圖 18 轉換數分組後的 b to a 數列.....	47
圖 19 A 與 B 實際的數列.....	47

圖 21 實驗二轉換次數 Box-plot	50
圖 20 實驗二轉換次數 Box-plot	50
圖 22 實驗三轉換次數 Box-plot	51
圖 23 實驗四轉換次數 Box-plot	51
圖 24 實驗一選擇 A 次數 Box-plot	52
圖 25 實驗二選擇 A 次數 Box-plot	52
圖 26 實驗三選擇 A 次數 Box-plot	53
圖 27 實驗四選擇 A 次數 Box-plot	53
圖 28 實驗一報酬 Box-plot	54
圖 29 實驗二報酬 Box-plot	54
圖 30 實驗三報酬 Box-plot	55
圖 31 實驗四報酬 Box-plot	55

表目錄

表 1 使用之報酬表.....	20
表 2 Goodnow 的實驗結果.....	22
表 3 報酬 Wilcoxon 等級和檢定.....	28
表 4 選擇 a 比例 Wilcoxon 等級和檢定.....	29
表 5 轉換次數 Wilcoxon 等級和檢定.....	31
表 6 各實驗的 b to a 的截斷點減去 a to b 截斷點 t 檢定.....	33
表 7 選擇 a 次數等級和檢定.....	35
表 8 平均報酬 Wilcoxon 等級和檢定.....	36
表 9 高轉換組選擇 A 比例和過去 10 選項 A 累積中獎次數的簡單 回歸模型.....	36
表 10 低轉換組選擇 A 比例和過去 10 選項 A 累積中獎次數的簡單 回歸模型.....	36
表 11 高轉換組與低轉換組截斷點 t 檢定	38
表 12 高轉換組選擇 A 比例和過去 10 到 1 次嘗試選項 A 累積中獎 次數的簡單回歸模型結果.....	48
表 13 低轉換組選擇 A 比例和過去 10 到 1 次嘗試選項 A 累積中獎 次數的簡單回歸模型結果.....	49

一、緒論

1.1 研究背景與動機

經濟學是一門研究如何選擇的學問，個體經濟理論對於消費者選擇模型強調在預算限制的條件下追求個人偏好極大化，這樣的模型在複雜的問題上，忽略作極大化選擇的困難，所以這樣的假設往往與現實的狀況不相符，現實因為環境等等的因素，使得一般人在做出選擇時難與理論模型完全相符。舉例來說，人在估計機率時會犯下系統性的錯誤 (Camerer, 1987 ; Tversky and Kahneman, 1974) 、違反效用函數遞移性公理 (Tversky, 1969) ，類似系統性的偏誤都是個體經濟理論對於人的行為假設的不足之處。

人們面對不確定的商品與服務時是如何做出選擇？日常生活中的選擇，像是購買商品或是服務，在購買前不會知道關於商品的特性（例如機率分配、不確定程度），但我們可以透過多次的購買來瞭解商品的本质。儘管現實狀況頗為複雜，許多選擇的問題的核心都可以抽絲剝繭成“N-arm bandit problem”，這個問題的假設是選擇是一個連續多期的行為，關於商品的特性為不確定，且資訊只能從選擇後的反饋獲得。

N-arm bandit problem 指的是一個決策者面對不確定的選項，只能從嘗試的過程中獲得關於選項的資訊，如同在賭場裡面對兩台長像一樣的吃角子老虎機，雖然賭場可能會告訴你中獎機率沒有差異，但你知道想要贏得大獎 (Jackpot) 那就要從中挑出一個有較高中獎機率的機台，在你以前沒有玩過這兩臺吃角子老虎機前，顯然你只能以不斷拉下把手的方式來判斷哪台的中獎機率較大。感覺起來賭博不是一個常見的例子，但事實上每天的生活很多時候我們都在選擇不確定的選項，像是選擇新香味的洗髮精、選擇自己習慣的洗髮精、新款式的智慧型手機或是速食店推出新餐點並且贈送玩具，這些從未消費過的商品，你不知道使用帶來效果如何前該如何選擇？N-arm bandit problem 問題所陳述的就是一個類似的

選擇問題。因為 N-arm bandit problem 能反映出日常生活選擇核心的部份，此研究我們試著利用較簡易的版本 two-armed bandit。

Two-armed bandit 問題是一個二擇一的問題，在這個問題上個體所面對的問題不外乎是開發 (exploitation) 與探索 (exploration) 行為間選擇的兩難，開發指的是探索選項不確定性質，探索則是待在目前停留的選項上充分享受這個選項所帶來的好處。開發與探索各有其優缺點，假設個體現在面對兩個不確定選項，其中一個有高報酬另一個有低報酬，個體在探索時會因為不停留在同一個選項錯過了高報酬選項可帶來的好處，相對的個體選擇開發時則是可能因為停留帶來低報酬的選項而帶來損失。

1.2 研究問題

過去研究關於 two-armed bandit 的文獻，如果個人是使用統計的方法來做出決定，這個問題的最適解應該是以貝氏理論為基礎的最適理論模型，如 Gittins 指標需要的計算較為複雜¹，除了當前與過去的結果、同時還要考慮選擇結束的時間、計算期望值等，如此最適化的困難，我們很難相信這樣的模型在預測選擇結果的預測能力。

其他過去的文獻，有的使用簡單的準則如截斷點策略 (cut point)²，以反映個人有限的計算能力，相對的這個準則所需門檻低—只需要簡單的記憶過去幾次的結果，這樣的方法在解釋實驗的資料相比也有不錯的結果，但不能明確的指出受測者在實驗時是使用這項準則，或進行相關的計算過程。

¹ 請參閱 J. C. Gittins and D. M. Jones. "A Dynamic Allocation Index for the Sequential Design of Experiments," Progress in Statistics, 1974, pp. 241-266.

² 請參閱 Banks, J., Olson, M. and Porter, D. "An experimental analysis of the bandit problem," Economic Theory, 1997, 10, pp 55-77.

本研究以兩個選擇的“arm bandit problem”版本，以實驗的方式模擬人在面對不確定選擇下的行爲。過去類似的實驗曾由 Jacqueline Jarrett Goodnow (1955) 執行過，他們得到的結果是如果兩個選項機率相差懸殊很大，受測者可以較快的收斂到機率高的選項，相反如果機率很接近時受測者在受斂到高機率選項的速度較慢。我們設計的實驗特別的著重在比較相同的機率上受測者接收到不同類型的反饋，來探討對不確定學習的過程、受測者表現與受測者對報酬“0”出現的敏感程度。

1.3 研究目的

過去學界有針對“arm bandit problem”研究，部分是以理論的方式來推導選擇的最適解，在用其所建立的模型與實驗的結果比較。本文以 two-armed bandit 爲架構，並且比較只提供當期結果、過去表現的資訊回饋和過去行爲的資訊回饋受測者實驗裡表現的差異，因此本文研究之目的及主要探討內容可簡單歸納如下：

1. 研究受測者學習的速度，不同程度的資訊是否可以讓受測者在實驗的初期就表現更好。
2. 受測者的選擇的型態，選擇探索兩個選項的狀況亦或是停留在熟悉的項目，這兩種行爲與反饋的關係。
3. 探討不同反饋下截斷點策略 (cut point) 的差異，是否不同的資訊回饋會影響個體對當前所選項目的信心？
4. 以四個實驗所有受測者轉換次數的中位數進行分類，來研究是否個體擁有自己決策的方式而忽略我們給的任何資訊。

二、. 文獻回顧

2.1 Two-armed bandit 實驗

人們重覆的對兩個完全不清楚的選項選擇，來探索兩個選項不確定的特性過程，個人會偏好哪個選項?除了會受到報酬的影響外，同時也會受到不確定程度的影響。(Einhorn and Hogarth, 1985) 在現實的世界裡，從不同的商品中做選擇，通常都是重覆的行為，消費者有誘因做出策略性的選擇，不只考慮當下選擇立刻帶來的效用也考慮後續選擇的效用。

Herrnstein et al. (1993) 實驗設計是受測者在電腦畫面上左右兩個送料斗 (hopper) 中選擇，當受測者選擇了一個某個送料斗後掉下錢幣表示獲得報酬，同時下方會顯示受測者所累積的報酬，使用的報酬函數特性是依照受測者過去左 (右) 邊送料斗被選擇的比例來決定本次左 (右) 兩個送料斗的報酬。作者根據報酬函數特性 (例：左右邊的報酬曲線是否交叉、線性的報酬函數) 與報酬定義 (例：掉下錢幣的間隔時間的長短、掉下錢幣表示損失) 的不同進行實驗。

實驗進行時受測者在兩個選項間做選擇，而這兩個選項獲得報酬機率為 1.0，但是能得到報酬不同 (例：報酬的數量不同、做完選擇與獲得報酬間的時間間隔長短不同) 根據受測者前幾次的選擇來決定，這種報酬會因為選擇改變的特性作者稱為內生性 (internalities)。作者依照選擇行為的不同定義出改善 (melioration) 與極大化 (maximization) 的行為，改善行為的特色是每次做決定時決策者會選兩選項間有較高報酬選項，極大化行為則是會考慮報酬的內生性，一個極大化決策者做決定時總是選能極大化總報酬的選項。Herrnstein et al. 的文章包含了五種不同的實驗，實驗一與二分別使用了交叉和平行的報酬函數，報酬的數量會隨著選擇變化，受測者有 150 次練習的嘗試次數與 400 次的實驗的嘗試次數，實驗三、四、五則是受測者作出選擇與報酬出現的間隔時間會隨受測者的

選擇變化，受測者有 300 秒的練習時間與 900 秒的時間進行實驗，實驗四針對兩種不同意義的報酬進行實驗（畫面上落下的錢幣代表贏得、畫面上落下的錢幣代表損失），實驗五則是在進行實驗前給受測者具有不同程度的指導語。

作者發現當實驗使用平行的報酬函數並且是報酬數量改變時，此時的反饋使受測者注意當前的報酬，作者認為如果資訊是有助於受測者計算當前兩個選項的報酬間的差異時，受測者在這樣的認知下會有改善的行為。

Herrnstein et al. 認為除了無法瞭解跟忽略資訊等認知的因素會造成偏離了最適化外，另一種可能原因是符合動機上的因素。像是短視折現 (myopic discount) 造成人們把立即的好處看的太過重要，相對的比較不考慮對往後效用的影響³。

在有內生性的環境下要做出極大化行為，需要理解當前兩個選項間報酬的差異之外同時也要考慮到內生性的影響。實驗裡受測者極大化行為同時受限於對報酬內生性的認知能力與當前兩個報酬間差異的認知能力，我們研究所考慮的方向是排除對報酬內生性認知能力的限制下個體不確定下的行為，如果在這樣相對簡單的狀況下受測者沒有出現極大化行為那在內生性下沒有極大化是不意外的結果。

Richard J. Tunney and David R. Shanks (2002) 同樣是考慮在有內生性的環境下並且增加 Herrnstein et al. 沒考慮的因素如：更多的期數、獲得報酬的機率不為 1.00 與獲得報酬機率不為 1.00 且具有負報酬的情況⁴。

Tunney et al. 所提出的的實驗架構與 Herrnstein et al. 相似，同樣的也是受測者在兩選項間選擇一個，和 Herrnstein 不同點是在於機率性的報酬函數，受測

³ 短視折現是指人在衡量跨期的效用時，每期折現率越後期的效用折現率越高。

⁴ 平均視窗數是本次報酬會受到幾期以前選擇結果的影響。

者每次的選擇不一定會有報酬，受測者面臨的困難是就算是做出最適化的選擇獲得的反饋也可能是“0”報酬。共有四種不同的設計，受測者都有 1000 次的嘗試次數並且使用平行的報酬函數，第一個設計：選擇結果改變的是報酬的數量；第二種設計：選擇結果改變的是得到報酬的機率；第三與四種設計：除了獲得報酬機率改變同時受測者有可能面對負報酬，另外第四種設計在實驗前給受測者額外 100 次的練習次數。當回饋是改變數量的狀況出現時，受測者出現極大化行為是高於改變機率的狀況，同時作者認為如果受測者有無處罰的機會學習選擇與報酬的影響能避免出現改善的行為。

Tunney et al. 把內生性延伸到獲得報酬機率的改變，在這樣的狀況下個體要知道報酬間的關係需要更多的時間。此篇所提出獲得報酬機率的改變更進一步的提高受測者做出極大化行為的困難，同樣的受測者因為無法了解報酬的內生性而有無法做出極大化行為。以上兩篇文獻皆是在有內生性的環境下說明個體行為偏離最大化的狀況，但內生性的情況無法說明一般的狀況下個體面對不確定資訊反饋與認知的影響，因此我們的設計排除了選擇的內生性，採用兩選項產生報酬機率固定的方式考慮對受測者選擇的影響。

Jacqueline Jarrett Goodnow (1955) 實驗設計則相對的比較簡單，這個實驗裡受測者需要在兩個按鈕中選擇一個，受測者是否贏得報酬是根據按鈕的機率所決定，報酬的機率分別是 100 : 0、50 : 0、100 : 50 和 75 : 25，除了 100 : 0 的組別實驗的嘗試次數是 75 次其他的組別嘗試次數皆是 150，在機率為 100 : 0、50 : 0 下作者並且比較了嘗試需要付費和免費兩種程序，Goodnow 的結果之一是當嘗試需要付費時收斂的速度比較快，另外機率為 100 與 50 時受測者在實驗結束時選擇可以收斂到較好的項目上，但機率為 50 與 0 時受測者即使在實驗進行到 140 次時仍有部分受測者選擇機率為 0 的按鈕。這裡獲得報酬的機率是較為簡單的狀況且內生性存在，機率不會隨著選擇的狀況改變 Goodnow 實

驗的實驗設計幫助我們建構一個簡單的環境下研究個體在不確定下的選擇問題，如果在較為簡單的狀況下人們不是做極大化行為，那這也可以解釋在複雜的情況下個體偏離極大化的情形，從 Goodnow 的實驗裡機率 0 與 100 兩種確定的狀況有不一樣的效果，我們認為人不是完全使用機率做決策，可能還考慮到其他資訊。

日常生活人們在做決策時，很多情況下是根據過去的經驗來當做參考，這種決策法則的特性是不需要花很多時間思考，儘管這不是一個理性的決策過程，最後的結果不是極大化效用，但優點是能很快速的做出決定。這樣避免了為做出理性的選擇花了過多時間計算的問題，Daniel Read, Loewenstein, and G., Kalyanaraman, S. (1999) 研究 Two-armed bandit 問題中經驗法則的影響，結合了立即效果與多樣化 (diversification) 兩種經驗法則 (heuristic)⁵，作者的兩個實驗納入決策的時間點與得到報酬的時間點。作者設計了兩種實驗情境，第一種為受測者在混雜娛樂和經典電影的名單中選擇觀賞的片子，經典的電影相對於娛樂性值不高但較有教育意義或帶來文化的涵養，因此短期的效用不高但經典片從長期來看效用較佳。帶有娛樂性質的片子較為有趣但也很容易被忘記，因此娛樂片子帶來的短期效用高但長期效用不如經典電影；另一種實驗的情境為選擇可立即對獎的彩票或是得獎與否需要幾天後才公布的彩票，立即對獎的彩票有立即且短暫效用，延後對獎的彩票則是因為持有彩券的期間會預期它中獎，而帶來較有持續性的效用。受測者在以做決定的時間點和選擇的商品（電影或彩票）消費的時間點安排成四組進行實驗。決定的時間點分為立即（在第一期決定接下來三期所消費品）與一系列（第一期決定的一期消費物品，第二期決定第二期消費物品），

⁵ diversification heuristic 是一種選擇上的經驗法則，人們同步的做出決策時（例：現在選擇 6 個點心在接下來的三個裡禮拜消費），人會容易尋求多樣性（例：選擇更多不同種類的點心）。

商品消費的時間點分為立即（第一期做完決定後能立即取得第一期商品）與遞延（做完決定後需等待至下期才能取得商品）。

作者透過以上這兩個實驗證實，第一、在立即的狀況下決定消費的商品與商品消費遞延的狀況，因為誇張的時間折現（Hyperbolic time discount）的關係個人偏好選擇消費效用需要長時間才能完整享受到的商品⁶。第二、立即做出決策與一系列做出選擇的兩種狀況比較下，因為多樣化的關係前者選擇的商品較會混合兩種不同屬性的商品。

Fernie, G., Tunney, R.J. (2006) 使用的愛華賭博實驗 (Iowa Gambling Task) 是一個比 two-armed bandit 複雜的問題，除了選項間離散的程度不同外甚至可能有負報酬出現。愛華賭博實驗被廣泛的運用在實驗上，用來測試真實生活的決策行為，然而根據研究者不同的設定受測者會有不同的表現，因而得到不同結論。綜觀過去的實驗進行實驗有的是使用代幣來當作強化物 (reinforcer)，其他的實驗者使用真錢，更重要的受測者所收到的指示也不同。Fernie et al. 討論這兩者間的差異，並且先前的研究對同一種強化物沒有使用兩種不同版本的指示。作者研究的目的是在比較兩種不同的強化物與兩種不同的指示，對愛華賭博實驗表現的影響。

愛荷華賭博實驗，實驗的參與者需要從四堆的撲克牌堆裡重複的選取牌出，每組撲克牌堆有不同報酬期望值，也有不同的賞金與損失出現的順序，裡面會有兩組牌堆有能帶來立刻的高賞金特性，但長期的期望值是負的（不利的牌堆，A 與 B 牌堆），另外兩組牌堆立即的賞金較低，但長期的期望值為正（有利的牌堆，C 與 D），這些牌堆出現損失的頻率也不同，A 與 C 牌堆有經常的損失（十次的選擇裡會有五次損失），B 與 D 牌堆出現損失頻率較低（十次選擇

⁶ (Hyperbolic discount) 函數： $u(x_0, x_1, \dots, x_t, \dots) = u(x_0) + \beta \sum_{t=1,2,\dots} \delta^t u(x_t)$. 第 0,1,2,3,... 期報酬的折現因子 $1, \beta\delta, \beta\delta^2, \beta\delta^3, \dots$ 。

會有一次損失)，IGT 表現的衡量方法是選取有利牌堆的次數減去選擇不利牌堆的次數。

實驗的第一場強化物在受測者學習的效果只有在受測者沒有指示的情況下才會出現，但是到第二場實驗時強化物的效果就不明顯了，顯示強化物對學習的影響不穩定。暗示程度較低的指示受測者學習的跡象不明顯，具有暗示性質該怎麼做的指導語的確可以幫助受測者表現的更佳。

以上幾篇以 two-armed bandit 問題為基礎的文獻，研究受測者能在什麼樣的反饋或環境下能瞭解兩個選項的特性，以實驗的表現來比較學習的速度，認為個人和最適選擇產生偏離的原因是認知與經驗法則的關係 (Herrnstein 1993)。除了反饋與實驗環境會影響選擇的過程外，實驗的指導語和實驗使用的強化物 (reinforcer) 也會有影響 two-armed bandit 上的選擇行為。

2.2 Two-armed bandit 實驗結果與模型的對照

以下的兩篇文獻以提出模型的方式來研究個體在 two-armed bandit 下的行為，他們所提出的模型皆考慮到了個體的認知能力，與本研究試圖從資訊的反饋中研究個體的行為的想法相同。

Gittins 指標是假設 bandit problem 具有暫時定態下決策者會有最適策略，而這樣的策略是以決策者對不同可選項目的主觀機率為基礎，在有限次的 bandit problem 下每一個選項可能的型式，決策者可以根據每個選項的特性分配一個數字，這樣最適化的過程就簡化成選擇有最大數字的選項。這種指數化策略，強調了決策問題下“解”的安定性，也使得比較靜態容易計算。例如改變與選項相關的機率分配，或改變對可選項目的事前機率我們可以推導出這些因素改變如何影響指數和影響指數化的策略。最適策略的安定性和特定的比較靜態假設。

Banks, J., Olson, M., and Porter D. (1997) 執行兩個 bandit problem 實驗討論幾個會影響比較靜態結果的因素，討論兩種 two-armed bandit 問題下的截斷點策略是否相同，風險偏好、機率、折舊率對截斷點的影響。第一個 two-armed bandit 問題的兩個選項，其中一個選項固定支付報酬“0.50”；然而另一選項報酬是不確定的，有可能是優或劣兩種型態，優的選項型態產生報酬“1” 機率為“g”，劣的選項產生報酬“1” 機率為“n”，且 $g > n$ 選項型態為“優”的期望值高於選項型態為“劣”。第二個 two-armed bandit 問題兩個選項支付的報酬都是不確定的都有可能是優或劣兩種型態。

Banks et al. 使用 Becker, DeGrootley, and Marschak (1964), BDM 的步驟分出受測者特定的風險態度，使用這個步驟可以歸納出實驗對象中特定的風險態度。受測者的電腦畫面上會顯示一個輪盤，這個輪盤上包含有 20 個數字，如果輪盤轉動後指向某幾個特定的數字，受測者將不會得到任何的報酬；如果沒有指向特定的數字，受測者將會得到報酬。受測者首先會被要求向這個輪盤的賭局出價，接著輪盤將會轉動，如果輪盤轉動的結果高於受測者的出價，受測者可以得到他對賭局的出價；如果輪盤轉動後的結果低於受測者出價，則接下來進行上述輪盤賭局。實驗進行時受測者會先經過 BDM 的步驟後在進行兩種 two-armed bandit 問題。在開始進行 two-armed bandit 選擇時，會先從 g (選項得獎機率) 與 δ (本回合是否繼續機率) 的組合裡抽出一組，在每次選擇結束時畫面上會有一個輪盤決定下次是否還是使用相同的機率，如果下次不繼續使用相同的機率則本回合結束並重新抽出一個機率組合並開始新的回合⁷。

⁷ 使用的機率組合有： $\delta = 0.8$ $g = 0.7$ 、 $\delta = 0.8$ $g = 0.9$ 、 $\delta = 0.9$ $g = 0.7$ 、 $\delta = 0.9$ $g = 0.9$ 。

這兩種 two-armed bandit 問題的最適選擇是，決策者根據先前的選擇更新信念，來考慮不確定的選項是好或壞，貝氏更新信念可以用以下表示：

對第一種 two-armed bandit 而言：

$$C(B_0) = 0$$

$$C(B_t) = \begin{cases} C(B_{t-1}) + 1, & \text{如果在 } t-1 \text{ 時選擇 B 並且結果是高報酬} \\ C(B_{t-1}) - 1, & \text{如果在 } t-1 \text{ 時選擇 B 並且結果是低報酬} \\ C(B_{t-1}), & \text{如果在 } t-1 \text{ 時選擇 A} \end{cases}$$

對第二種 two-armed bandit 而言

$$C(B_t) = \begin{cases} C(B_{t-1}) - 1, & \text{如果在 } t-1 \text{ 時選擇 A 並且結果是高報酬} \\ & \text{或在 } t-1 \text{ 時選擇 B 並且結果是低報酬} \\ C(B_{t-1}) + 1, & \text{如果在 } t-1 \text{ 時選擇 A 並且結果是低報酬} \\ & \text{或在 } t-1 \text{ 時選擇 B 並且結果是高報酬} \end{cases}$$

使用 $C(B)$ 可以建立一個臨界值信念截斷點 (critical belief cut point) 策略。

對第一種 two-armed bandit 而言

$$c \in \{ \dots, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2 \dots \}$$

$$\sigma_t = \begin{cases} B, & \text{如果 } C(B_{t-1}) > c \\ A, & \text{如果 } C(B_{t-1}) < c \\ A \text{ or } B, & \text{如果 } C(B_{t-1}) = c \end{cases}$$

對第二種 two-armed bandit 而言

$$\sigma_t = \begin{cases} B, & \text{如果 } C(B_{t-1}) > 0 \\ A, & \text{如果 } C(B_{t-1}) < 0 \\ A \text{ or } B, & \text{如果 } C(B_{t-1}) = 0 \end{cases}$$

作者同時也檢驗截斷點策略是否能描述決策的行為，實驗的結果顯示兩個不同種類的 bandit problem 裡用的截斷點策略是不同的，第二種 two-armed bandit 裡截斷點行為是與近視的行為一致，但在問題較為簡單第一種 two-armed bandit 裡卻不是。同時第二種 two-armed bandit 裡的截斷點策略不會受到風險態

度、折現率、事前機率、報酬的影響，其中作者提到了截斷點策略本身並不是一個穩定的策略。前面提到截斷點策略不會因為受測者各別風險態度的改變，隱含著受測者在決策時不是使用極大化的策略，我們認為受測者在面對不確定的問題比較可能是使用與風險態度無關的策略。Grether (1992): “在不確定環境下不同狀況下個體使用不同的決策法則” 而身為經濟學家的我們想要發現的是從決定策略方面來說穩定的變數或因素。

Gans, N., Croson, R., and Knox, G. (2004) 則是以抽撲克牌的情境來進行實驗，他們文章的主要研究方向為比較幾個模型在解釋受測者的行為上的解釋能力。實驗對象可以從兩堆的撲克牌裡選擇，每個牌堆是由 “YOU WIN!” 與 “YOU LOSE!” 兩種卡組成，每次實驗對象選一個牌堆，電腦會隨機的從那個牌堆裡抽出一張牌，在展示贏或輸過後抽出的牌會在放回原來的牌堆裡；牌堆的相對機率分別為 “0.4 對 0.4” 、 “0.4 對 0.15” 和 “0.40 對 0.65” 。實驗對象可以知道他們累積的報酬、累積的 win 卡與 lose 卡與過去歷史的選擇與選擇結果。

Gans et al. 進一步測試，可行且合理的模型來解釋 two-armed bandit 裡的選擇行為，作者使用三種相關的貝氏模型和三種有限制記憶空間的模型，解釋一系列的 two-armed bandit 問題下人如何做出選擇，這裡並且分別使用三種方式評估貝氏模型的表現。

第一，檢驗三種不同貝氏模型裡有一致結果的預測：預期實驗對象會在某個選項上連續嘗試的次數，預期實驗對象採取轉換行為的時間與和選項報酬是增加且凸 (increasing and convex) 的關係。第二，檢驗每一個模型預測的實驗對象每次的選擇行為。最後比較三種貝氏模型和三種非貝氏模型的配適程度。Gans et al. (2004) 文章裡所使用屬於貝氏的模型：

1. Gittins-Index Model

在 t 時選擇 i 選項可以得到效用 U^i ， U^i 的是一個隨機的分配，每一個 U^i 是從 Θ 裡抽出，每一個 U^i 可以用 θ 表示， $U^i \sim F(u|\theta^i)$ ， $\mu_\theta = [U|\theta] = \int u dF(u|\theta)$ 。對每個選項而言，有一個累積分配函數 $P_t^i(\theta)$ ，表示在 t 時決策者對選項品質的瞭解程度，令 $\{P_0^1, \dots, P_0^m\}$ 為事先決策者還沒開始做選擇時的資訊，如果決策者在 t 時選擇 i 選項，得到效用 u 那他對 i 新的累積分配函數是：

$$dP_t^i(\theta|u) = \frac{dP_{t-1}^i(\theta)df(u|\theta)}{\int_0^\infty dP_{t-1}^i(\theta)dF(u|\theta)} \quad \forall \theta \in \Theta$$

如果沒有選擇 i 項: $dP_t^i(\theta) = dP_{t-1}^i \quad \forall \theta \in \Theta$

令決策集合 $\pi = \{\pi(1), \pi(2) \dots\}$ 是一系列的選擇， Π 是非預期未來效用的決策集合，理性決策者從決策集合當中，尋找一種決策能夠極大化折現後的未來價值

$\text{Sup}_{\pi \in \Pi} E_\pi[\sum_{t=1}^\infty \alpha^t U_t]$ ， α 為折現率。Gittins 指標是每一單位預期折現時間下的預期折現效用：

$$G \left(P_t^j \triangleq \text{Sup}_{T>\tau} \left\{ \frac{E[E[\sum_{s=t+1}^{T-1} \alpha^{s-t} U(P_{s-1}^j | P_{s-1}^j)]]}{E[E[\sum_{s=t+1}^{T-1} \alpha^{s-t}] P_{s-1}^j]} \right\} \right)$$

2. Myopic Model

一個短視的決策者不用去考慮結束時間，只要簡單的決定較喜歡那個選項

$$M(P_t^j) \triangleq \frac{\sum_{s=t+2}^\infty \alpha^{s-t}}{\sum_{s=t+1}^\infty \alpha^{s-t}} = E[U(P_t^j)] = \int \mu_\theta dP_t^j(\theta)$$

他會選擇 i 選項來極大化長期效用: $i = \arg \max \{M(P_t^i)\} \{M(P_t^j)\}$

3. Simple Model

一個簡單的決策者，他把可能的品質為兩個分類—好與壞，效用分配函數為 $F_G \equiv F(u|\theta^G)$ 與 $F_B \equiv F(u|\theta^B)$ ，簡化選項可能的品質分配為 $\Phi = \{\theta^B, \theta^G\}$

令 $\mu_G \triangleq E[U|\theta^G]$ ， $\mu_B \triangleq E[U|\theta^B]$ ，簡單決策者在計算 $M(P_t^j)$ 是用：

$$S(P_t^i) = E[U(P_t^i)] = \mu_G P_t^i + \mu_B (1 - P_t^i)$$

$S(P_t^i)$ 是預期在期選擇 i 選項效用，決策者用貝氏法則更新他們的事前機率可縮減為：

$$P_t^i = \frac{P_{t-1}^i dF_G(u)}{P_{t-1}^i dF_G(u) + (1 - P_{t-1}^i) dF_B(u)} = \left(1 + \frac{1 - P_{t-1}^i}{P_{t-1}^i} \cdot \frac{dF_B(u)}{dF_G(u)} \right)^{-1}$$

屬於非貝氏模型：

1. Last-n

這個法則是爲了要去符合人記憶能力的限制，只考慮過去在每個選項上嘗試的 n 次。

2. Hot Hand

第二個法則是考慮了選擇的過程中，停留在某個選項上的行爲，實驗對象會停留在原來的選項上如果目前選的選項結果是“贏”，如果有 n 個“輸”出現則會轉換到另一個選項。

3. Exponentially Smoothed

這是一個指數的學習模型，受測對象以前次的報酬來更新他的記憶，在 t 時對選項 i 的指數爲：

$$ES_i(t) = (1 - \omega)ES_i(t - 1) + \omega I(\text{arm } i \text{ wins at } t) \quad , \quad \text{where } 0 < \omega < 1$$

使用 Gittins 指標建立的最適策略在計算上會非常困難，與實際上人們如何解決 two-armed bandit 問題可能有所出入，先前的研究結果指出人的行爲大約和貝氏模型預測一致，但會有嚴重且系統性的偏離，Meyer and Shi (1995) 發現實驗對象選擇反映貝氏的更新事前機率 (Bayesian updating of priors)，不過行爲會比最適策略所預測還要近視 (myopic)。

實驗的結果顯示選項的機率愈高，連續選擇那個選項次數愈高，在利用模型分析實驗對象的選擇方面，Simple Model 所產生的錯誤也低於其他的貝氏模型，Simple Model 整體的 BIC 指數低於其他的貝氏模型，當 Simple Model 與其他的非貝氏模型比較時，Simple Model 與 Myopic-1 在 BIC 指數低於 350 時表現比其他非貝氏模型較好，HH-5 與 HH-N 在其他的 BIC 指數範圍裡表現優於 Simple Model 與 Myopic-1。

前面 Banks et al. 文章使用了截斷點一個臨界值的概念分析 two-armed bandit 的行為，當兩個選項都是不確定時他們觀察到截斷點策略是與近視行為一致。另外 Gans et al. 比較模型間的結果在大部份的狀況 HH-5 與 HH-N 的模型比近視模型解釋能力好。

考慮到大部分的時候 Hot hand 有較佳解釋能力所以我們用類似的方式來計算受測者轉換前的門檻值，在計算截斷點策略時只簡單計算在轉換前報酬“0”所出現的次數。不管是截斷點策略或是 Hot hand 都是衡量受測者對當前選項信心的一種方式，如果受測者對當前選項非常的有信心那門檻值也會較高。

2.3 Two-armed bandit 資訊的獲取與學習

Gabaix et al. (2006) 則是從認知的過程中切入，以做出決策的過程中決策者資訊的獲取狀況研究人在面對不確選項時的行為，作者以直接認知模型 (Direct Cognition Model) 分析實驗的資料並與理性的模型比較。

直接認知模型 (Direct Cognition Model)，假設代理人行為是近視的計算選項的價值並且以此選擇他們下一步的認知操作，本文用兩個實驗裡資訊的獲取來檢驗此模型。在第一個實驗裡資訊的獲取有明顯金錢上的成本，第二個實驗資訊的成本是因為做選擇的時間有限。

第一個實驗裡受測者要在三種投資計劃中選擇，受測者可以知道這三個投資計劃獲得報酬的機率及報酬的價值，而這三個計劃裡其中會有一個計劃獲得報

酬的機率是 100%。受測者在做出決定前，可以先對計劃進行調查的動作（調查的動作需要成本），如果某個計劃被調查完後，受測者可知道此計劃實際上是否可以得到報酬。

第二個實驗受測者需要在八個商品中選擇分別有九種屬性，屬性是以數字表示，商品最後可以得到的報酬是由這九種屬性的數字所累積加總，受測者的螢幕上會顯示 8×9 的矩陣，受測者在決定要消費那項貨物前，可以透過點選螢幕上的方塊來調查某項商品的屬性，當方塊被選擇調查時方塊的數字會顯示出來。

直接認知模型，是假設每次進行調查的動作，都把這個調查動作當作最後一次執行，令 S_t 是在 t 時已知的最佳選擇， X_i 是報酬的實際值，預期能從計畫 i 中的獲利為：

$$E[\max(X_i, S_t)] - S_t = E[(x_i - S_t)]^+$$

執行調查這項操作的成本是 C_i ，直接認知模型會對最高獲利的項目執行調查的操作：

$$\begin{aligned} G_i &\equiv E[(x_i - S_t)^+] - C_i \\ &= p_i(V_i - S_t) - C_i \end{aligned}$$

直接認知模型會調查或採用有最高 G 值的計劃，直接認知模型會從有高機率計劃開始調查。對每次的嘗試來說低機率的計劃有 P_i 的機率可以獲得 V_i ，高機率的計劃有 P_j 獲得 V_j 這兩者有以下的關係：

$$\begin{aligned} \frac{P_i V_i - C}{P_i} &> \frac{P_j V_j - c}{P_j} > 1 \\ P_i(V_i - 1) &< P_j(V_j - 1) \end{aligned}$$

直接認知模型能預測受測者選擇調查哪項計劃，而在預測最後受測者所執行得計劃上，直接認知模型與理性模型的結果不存在很大的差異，事實上當直接認知模型預測結果與理性模型結果不同時，直接認知模型的結果比較接近受測者的資料。

以上所提到的直接認知模型所考慮的資訊在第一個實驗中是選定調查的計劃得獎與否，受測者得到這樣的資訊後選項對受測者來說都是確定的，個體在這樣的實驗裡沒有學習不確定狀況的過程，第二個實驗裡受測者可認知是只有選項某部分的報酬，但因為在不同的情況下受測者可能會有不同的策略。我們加入的資訊回饋除了關於報酬資訊之外，還多考慮了受測者自己過去行為的資訊回饋，這樣我們可以觀察到受測者在不同的資訊回饋下是否採行不同策略。

總結一下我們在本章所介紹有關於 **two-armed bandit** 的重要概念，第一節的文獻的方向是比較不同報酬函數、實驗進行時所給的強化物和指導語這兩個實驗參數的影響，造成受測者表現不同的原因是因為認知的關係，如果實驗所設定的參數能讓個體較容易認知出極大化行為所需要的資訊，個體可以較容易做出極大化行為。第二與第三節提模型來解釋個在 **two-armed bandit** 下的行為，他們文章裡有較好表現的模型特色是，沒有考慮太多的期數（像是 **Hot hand**）。關於應該使用哪種的模型來解釋一般的狀況，我們在這邊認為一個較為合理的模型應該是要類似於第三節所提到的直接認知模型一樣，考慮到受測者實際認知的部分，因為要做出最適策略需要的計算對一般人來說不太容易。

如同以上所述考慮受測者實際的認知到的資訊是解釋個體在 **two-armed bandit** 問題下行為的基礎之一，又因為在解釋模型這方面需要有更嚴謹的討論之外也需要考慮不同參數下（兩選項機率、期數的長短），基於以上的考量本文目前想要解釋的問題是不同資訊回饋下對受測者在不確定選擇行為上之影響，雖然這是一個較為簡單的問題，但這是一個重要的問題並且能在日後建立模型時提供一點建議。

三、 實驗設計

3.1 實驗概述

實驗一提供的反饋設計是比照 Jacqueline Jarrett (1911) 的設計，受測者只能看到當期的報酬，在實驗二我們增加受測者所可以得到資訊（利潤、總利潤、總報酬與稟賦加總），如此我們希望可以從對照兩個實驗反饋下的選擇行為瞭解受測者的決策過程中不同的資訊回饋的影響、受測者的學習行為。

實驗的設計是透過一個相對較簡單二擇一的問題，模擬真實世界個體在面對不確定選項選擇過程與學習的過程。每個實驗裡，受測者需要在兩個選項“A”、“B”間重覆連續的做出 150 次選擇，這兩個選項是否會得到是不確定的，受測者選擇後的報酬是有可能是“0”或是“1”，A 選項得到報酬機率 70%，B 選項獲得報酬機率為 30%，受測者如果持續選擇 A 選項，累積的報酬會高於持續選擇 B 選項。

實驗前不會告知受測者關於兩個選項報酬的資訊，在正式實驗前也沒有提供受測者練習的次數，當受測者做出選擇後會告知本次選擇得到的報酬，受測者只能從我們提供的反饋或報酬瞭解兩個選項報酬的特性。

我們給受測者在實驗裡表現更好的誘因是課堂上的成績，這樣的誘因對受測者來說可能不如金錢上的高，不使用金錢當作誘因的原因一方面是經費的限制另一方面是金錢誘因不一定能促進受測者表現。Camerer and Hogarth (1999) 比較

了 74 個實驗的結果指出金錢的誘因不一定可以讓受測者表現更好，在賽局、喊價與風險的選擇上典型的結果是更多的誘因不會影響平均表現，但誘因讓受測者回應的變異程度下降。

3.2 實驗參數

我們依照給予反饋的不同和使用的報酬函數不同分成四個實驗，資訊回饋包含了有期數、報酬、利潤、總利潤、總利潤和稟賦加總、總共選擇 A 次數和總共選擇 B 次數。計算的方式如下：

- 1.期數的計算方式是以受測者每次嘗試當作一期。
- 2.報酬是根據報酬函數設定決定受測者當期是否有中獎，如果受測者有中獎是“1”，沒有則是“0”。
- 3.利潤是每次得到的報酬減去嘗試成本，這裡的嘗試成本設定為“0”，嘗試成本在這裡是沒有影響的。這裡放入沒有影響的資訊原因是為了往後進行實驗時的彈性，我們可以很容易的在程式裡面設定我們想要的數字，而不用重新設計程式。
- 4.稟賦設定為“100”。

我們在每次選擇完後的結果畫面，讓參與實驗二受測者看到利潤、總利潤、總報酬、與稟賦的資訊。參與實驗一的受測者則是在實驗前的說明告知稟賦的資訊。期數和報酬的資訊回饋性質上是屬於有關於 A、B 兩個選項不確定的訊息，如果受測者能充分的瞭解這部份資訊，受測者的表現則應該會更佳，利潤、總利潤和總利潤和稟賦加總性質上是屬於受測者過去表現所累積表現，這部份的資訊則是幫助受測者判斷自己做的好或不好，總共選擇 A 與總共選擇 B 的次數是屬於受測者本身過去行為的紀錄，最後這個部份的資訊只是提醒受測者在過去做了甚麼事，我們預期對受測者的影響不大。

實驗裡使用了兩種不同型態的報酬函數，第一種型態的報酬函數受測者得到報酬“0”或“1”是決定於他們所在的期數，受測者每期選擇得到的報酬根據報酬表 A、B 欄所相對應的數字，假設受測者在第二期時選擇了 A，根據報酬表第二期選擇 A 的報酬為“1”受測者獲得的報酬就為“1”。

第二種型態的報酬函數受測者得到報酬“0”或“1”是決定於他們選了 A (B) 多少次，計算受測者過去總共選擇 A 或 B 的次數，報酬的決定是根據報酬表 A、B 所選擇次數所相對應的數字，假設受測者總共選擇 A 十次，選擇 A 第十次的報酬就為“1”。

表二是兩個實驗使用的報酬表，受測者每一期選擇得到的報酬，是根據報酬表 A、B 欄所相對應的數字，A 欄的加總為“105”，B 欄為“45”。

表 1 使用之報酬表

期數	A	B	期數	A	B	期數	A	B	期數	A	B	期數	A	B
1	0	0	31	1	1	61	1	1	91	1	0	121	1	0
2	1	0	32	1	1	62	1	1	92	1	1	122	1	0
3	1	0	33	0	0	63	1	0	93	0	1	123	1	0
4	1	0	34	1	0	64	0	1	94	1	0	124	1	0
5	1	0	35	0	1	65	0	1	95	0	1	125	0	0
6	1	0	36	0	0	66	1	0	96	0	0	126	0	0
7	1	0	37	1	0	67	0	0	97	1	0	127	0	0
8	1	0	38	1	0	68	1	1	98	1	0	128	1	0
9	1	1	39	0	0	69	0	0	99	1	0	129	1	0
10	1	1	40	0	0	70	1	0	100	1	0	130	0	1
11	1	0	41	1	1	71	0	0	101	1	0	131	1	1
12	1	0	42	1	0	72	1	0	102	1	1	132	0	0
13	1	1	43	0	0	73	1	0	103	1	1	133	0	0
14	1	0	44	0	0	74	1	0	104	1	0	134	1	0
15	0	0	45	1	0	75	1	0	105	1	1	135	1	1
16	1	0	46	1	0	76	1	0	106	0	0	136	1	1
17	1	1	47	1	0	77	0	0	107	0	0	137	1	0
18	0	0	48	0	0	78	0	0	108	0	0	138	1	1
19	0	1	49	0	0	79	0	0	109	1	0	139	1	0
20	0	0	50	0	0	80	1	0	110	1	0	140	1	1
21	1	1	51	1	0	81	1	0	111	1	0	141	1	0
22	0	0	52	1	0	82	1	0	112	1	0	142	1	1
23	1	1	53	1	1	83	1	0	113	1	0	143	1	1
24	1	1	54	1	0	84	1	0	114	1	0	144	1	1
25	0	0	55	1	1	85	1	1	115	1	0	145	1	0
26	1	1	56	0	0	86	1	0	116	1	0	146	1	0
27	1	1	57	1	0	87	0	1	117	0	1	147	1	0
28	0	0	58	1	0	88	1	0	118	0	1	148	1	1
29	1	0	59	1	1	89	1	0	119	0	1	149	1	0
30	1	1	60	0	0	90	0	0	120	1	1	150	1	0

我們實驗裡所使用的資訊反饋與報酬函數分別是：

1. 實驗一使用的反饋是當次選擇的報酬、期數，使用的報酬函數是第一種型式。

2. 實驗二使用的反饋是當次選擇的報酬、期數、利潤、總利潤、總報酬、總利潤與稟賦加總這幾項資訊，使用的報酬函數是第一種型式。
3. 實驗三使用的資訊是當次選擇的報酬、期數、利潤、總利潤、總報酬、總利潤與稟賦加總、總共選擇 A 次數、總共選擇 B 次數，使用的報酬函數是第二種型式。
4. 實驗四使用的資訊是當次選擇的報酬、期數、利潤、總利潤、總報酬、總利潤與稟賦加總、總共選擇 A 次數、總共選擇 B 次數，使用的報酬函數是第一種型式。

3.2.1 實驗一

在實驗一裡面共有 39 名受測者，其中的 10 名受測者為政治大學學生在政治大學進行實驗，其他的受測者則是在東海大學學生在東海大學進行實驗，實驗一的政大受測者為大學部同學及研究生，科系包括經濟系、心理系（只有一位）、經濟所和應用物理所，修習課程為實驗經濟學；實驗一的東海受測者為經濟系大學部同學修習課程為行為經濟學。

實驗裡每名受測者進行 150 次的嘗試，每次選擇完後結果畫面上會顯示受測者當次選擇報酬已及目前所在的期數。受測者只能知道當期報酬是否為 1 與期數依此簡單的資訊回饋做出選擇，這個實驗與 Goodnow 實驗類似。

Goodnow 實驗中機率 0.75 對 0.25 的組別是以嘗試需要付費的條件 (pay to play)，在這個條件下支付給受測者的報酬是贏的部份與輸的部分相減條件下受測者表現不如其他的組別，這種結果有可能是因為機率的差異，也有可能是免費玩的與付費在第一種情況下受測者面對的報酬是“0”與“1”，第二種情況下是“1”與“-1”兩種情況下的期望值不同，由於在有損失的情況下個體的行爲有別於無損失的情況，我們實驗設計以免費玩的方式進行，實驗一我們所想要以報酬不會出現損失的情況來說明個體的決策行爲。

表 2 Goodnow 的實驗結果

Cost	100:0		50:0		100:50	
	Pay to Play	Play Free	Pay to Play	Play Free	Pay to Play	Pay to Play
0-9	0.54	0.38	0.44	0.44	0.54	0.42
10-19	0.84	0.66	0.70	0.50	0.54	0.68
20-29	1	0.90	0.60	0.70	0.64	0.54
30-39	1	0.88	0.70	0.68	0.66	0.78
40-49	1	0.90	0.70	0.58	0.84	0.76
50-59	1	0.88	0.82	0.74	0.96	0.68
60-69	1	0.99	0.86	0.60	0.90	0.66
70-79			0.90	0.68	1	0.80
80-89			0.88	0.76	1	0.80
90-99			0.80	0.70	1	0.88
100-109			0.94	0.76	1	0.92
110-119			0.78	0.68	1	0.90
120-129			0.78	0.84	1	0.92
130-139			0.82	0.76	1	0.82
140-149			0.90	0.74	1	0.86

3.2.2 實驗二

實驗二受測者共有 31 人其中的 10 名為政治大學學生剩下的 21 名受測者為東海大學學生，實驗二的政治大學受測者為博士和碩士學生系所為經濟系；實驗二的東海大學受測者為大學部同學，修習課程為行為經濟學。

同樣也是進行 150 次嘗試和與實驗一相同的報酬計劃表。實驗二與實驗一不同的是所使用的資訊回饋，實驗二裡所使用的資訊回饋比起實驗一增加了總報酬、總利潤與稟賦加總，這些資訊能反映出受測者在實驗裡的表現，受測者每次選擇時都可以知道自己目前累積的報酬數量，如果受測者決策時考慮的是極大化整體報酬這幾項資訊可以讓受測者判斷自己做的正確與否。實驗二受測者應該

能更快的增加選擇 A 次數，比實驗一選擇更快收斂至 A 上。另外雖然實驗一的受測者仍然可以透過自己的記憶加總每次的資訊，得到與實驗二相同的資訊，但在進行相同的決策過程時如果能在第一個步驟時就告訴受測者每次選擇的加總使他們不用額外的把結果轉換成數字，受測者能比較容易做出決定。

3.2.3 實驗三

在實驗三裡受測者 34 名受測者全部為東海大學學生，受測者為大學部學生，修習課程為個體經濟學。實驗三使用的報酬表則是第二種型式的，資訊回饋包括有報酬、期數、利潤、總利潤、總報酬、總利潤與稟賦加總、總共選擇 A 次數和總共選擇 B 次數。

實驗一與實驗二的實驗結果展現了兩種不同的資訊回饋受測者的表現找不出很大的差別，所以實驗三裡我們進一步給的資訊回饋除了包含實驗一與二的資訊回饋外，還導入了受測者自己過去行為的紀錄。實驗三裡的資訊回饋能提醒受測者他們自己過去行為，如果受測者決策的某個步驟是考慮到自己過去的選擇(例如過去選擇 A 次數、A 中獎的頻率)，那我們給的資訊回饋可以讓受測者很容易轉換成他們所需要的資訊。

3.2.4 實驗四

實驗四受測者有 57 位東海學生，受測者為大學部同學修習課程為經濟學，系所為統計系。實驗四使用的資訊是當次選擇的報酬、期數、利潤、總利潤、總報酬、總利潤與稟賦加總、總共選擇 A 次數、總共選擇 B 次數，使用的報酬函數是第一種型式。

實驗三裡受測者表現明顯的與實驗一、二不同，然而因為實驗三同時在報酬函數和資訊回饋上的設定與實驗一實驗二不同，這樣的差異是兩者共同影響的，爲了要瞭解資訊回饋在決策上的影響因此實驗四裡使用與實驗一、二相同的報酬計劃，排除不同報酬計劃的影響。實驗四的設計是假設受測者採用的是相對

次數的策略或類似的決策方法，那受測者在這個設計下會更容易選擇較好的選項。

3.3 使用軟體

我們使用 Z-tree 程式語言來設計我們的 bandit problem 實驗，Z-tree 軟體是由 Zurich 大學所研發，這套軟體的功能是用來設計並執行實驗，而受測者的介面則是 Z-leaf，Z-tree 程式的特色是：

- 1.Z-tree 程式能降低設計實驗程式的難度，只有一點寫程式知識的使用者能夠在幾個小時內創造出簡單的實驗。
- 2.使用 Z-tree 能很容易執行實驗，軟體內已經內建了幾個常見的經濟實驗會有的項目，例如受測者報酬。
- 3.Z-tree 很有彈性，不同的實驗能很容易的組織或組合在一起，並且能夠顯示圖形。
- 4.Z-tree 資料儲存上是安全的，所有的資料將會自動的儲存。

Z-tree 軟體已經廣泛的使用在實驗上，像是公共財實驗以及喊價的實驗。Z-tree 軟體能記錄受測者實驗過程中選擇的行為，與記錄受測者在實驗過程中獲得的資訊與報酬，這樣我們可以知道受測者每次做決定前所看到的資訊。

3.4 實驗流程

實驗共有 161 人參與這實驗，受測者為政治大學與東海大學學生，東海進行實驗的地點是資訊教室，受測者參與實驗可以得到課堂的分數做為參與實驗的獎勵，每人只進行一種實驗，其中的 27 人參與實驗一，21 人則是實驗二。

我們使用 Z-tree 軟體進行實驗，實驗前的指導語是透過電腦螢幕顯示給受測者，受測者在各自的電腦上面會顯示 A、B 兩個按鈕，受測者以滑鼠點擊按鈕來選擇，並且在按鈕點擊完後進入結果畫面，然後進行下一次 A、B 兩個

選項的選擇，實驗限制受測者需要在 60 秒內做出選擇，否則就會跳入下一期的選擇，沒有受測者選擇所花費時間超過我們的限制，實驗的流程如下：

- (a) 每個受測者使用一台電腦，電腦上的起始畫面是 Z-tree 軟體的等待畫面，在實驗開始前，在由主試者引導開啓指導語畫面。
- (b) 受測者在閱讀完指示按下開始鍵後，受測者開始進行實驗一或實驗二。
- (c) 受測者在畫面中選擇 A 或 B 按鈕後進入結果畫面，結果畫面會顯示我們給予受測者的反饋。
- (d) 受測者閱讀完我們給予的資訊後可以按下“OK” 鈕進行下次的選擇，重覆 150 次後實驗結束。

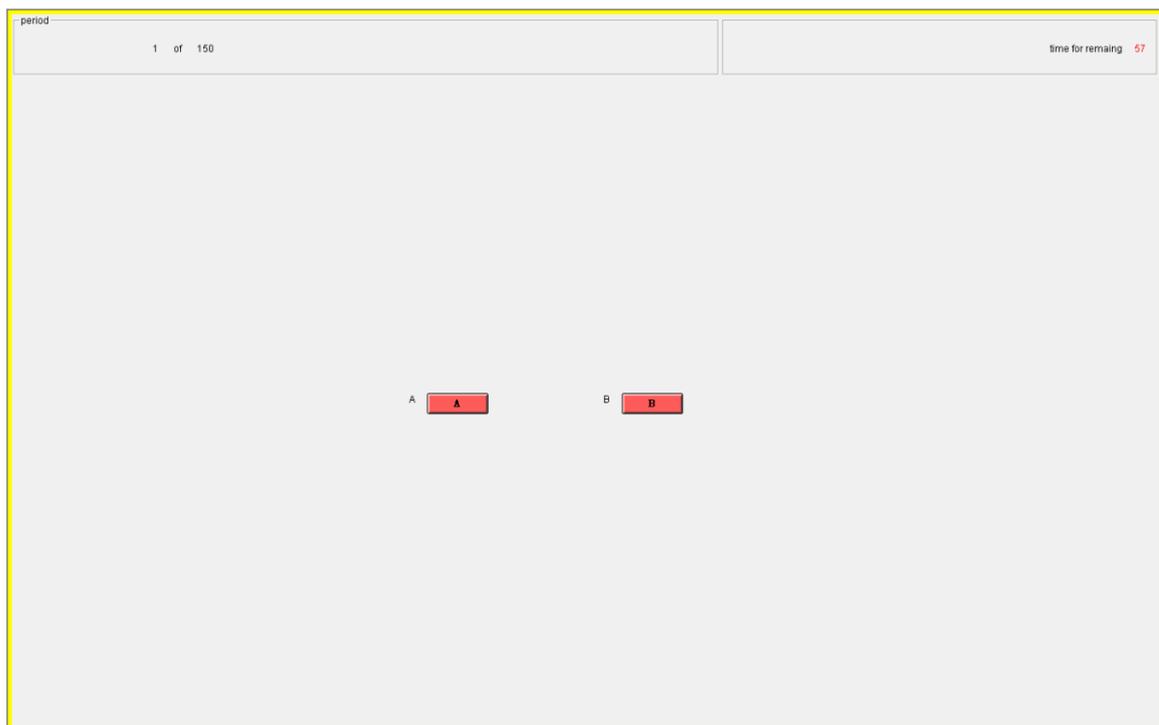


圖 1：實驗中選擇畫面



圖 2：實驗一 結果畫面

四、實驗結果

4.1 平均表現

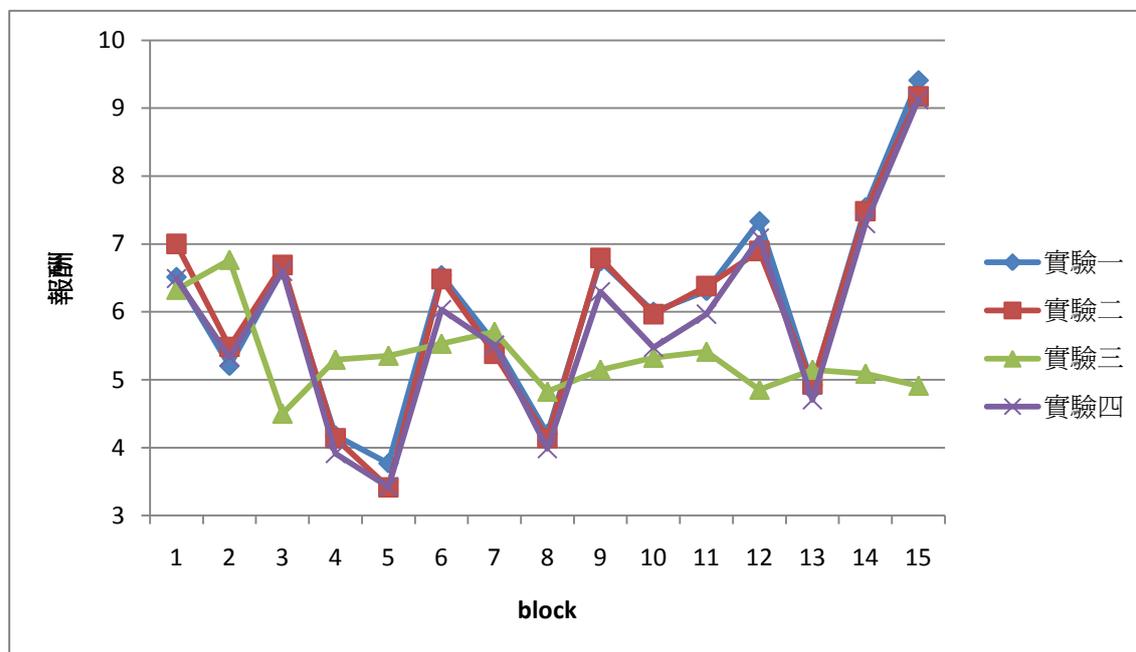


圖 3 實驗一到實驗四的平均報酬

每一個實驗的報酬是以每十次嘗試為一個 block 計算，報酬在每個 block 內的範圍是 0 到 10，圖 3 是四個實驗裡所有 block 上的平均報酬，實驗一整體的平均報酬是 6.05，實驗二整體的平均報酬是 6.02，實驗三整體的平均報酬是 5.35，實驗四整體的平均報酬則是 5.82。實驗三中位數 Wilcoxon 等級和檢定與其他實驗顯著的不同。圖 3 平均報酬變化的趨勢實驗一、實驗二與實驗四平均報酬變化的趨勢大致相同，實驗四是一個給予資訊回饋更多的狀況但在平均表現上與資訊回饋最少的實驗一差距不大，在我們的例子裡給予受測者不同的資訊回饋沒有影響對受測者報酬沒有幫助。

實驗一、二、三平均報酬最高點都出現在實驗結束的 block 上，實驗三的平均報酬並沒有隨著 block 數增加甚至最後一個 block 時出現向下的趨勢。

表 3 報酬 Wilcoxon 等級和檢定

平均報酬	z	雙尾 p 值	n
Exp1 vs Exp2	0.00665558	0.99469	150
Exp1 vs Exp3	4.73212	< 0.001**	150
Exp1 vs Exp4	0.851249	0.394631	150
Exp2 vs Exp3	4.70417	< 0.001**	150
Exp2 vs Exp4	0.823961	0.409962	150
Exp3 vs Exp4	-3.83362	0.000126273**	150

4.2 選擇的狀況

接著我們從受測者選擇多少 A 選項次數，觀察資訊回饋對受測者選擇行為的影響，圖 4 實驗一到實驗四選擇 A 的比例是四個實驗每個 block 上選 A 人數佔整體受測者比例。實驗二在 block 6 以前選擇 A 比例的趨勢高於實驗一，在 block 6 之後實驗一與實驗二開始出現一樣的趨勢。實驗一與實驗二的差距是不顯著的 (p=0.403188)。

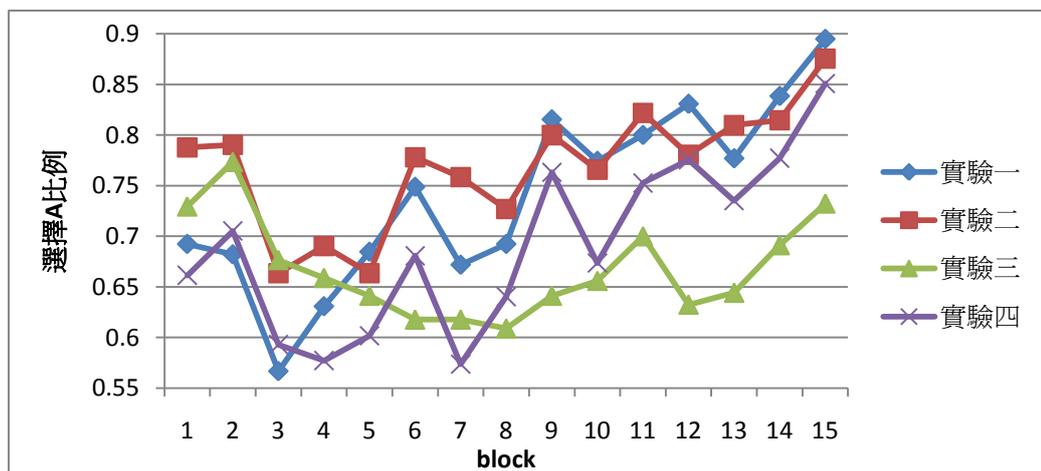


圖 4 實驗一到實驗四選擇 A 的比例

表 4 中實驗一與實驗四 Wilcoxon 等級和檢定是顯著的 ($p=0.000671391$)，實驗三與實驗四 Wilcoxon 等級和檢定同時也是顯著的 ($p=0.059809$)，表示實驗四選擇 A 次數是少於實驗一與實驗二的。

實驗四雖然選擇 A 的次數在所有的 block 裡都少於實驗一與二，但仍然和實驗一與二呈現相同的趨勢，顯示給予有關受測者本身過去的選擇資訊回饋反而會使受測者對較好的選項信心下降。所有的實驗裡選擇 A 的比率都在 0.5 以上表示受測者較偏好 A 選項，這隱含著受測者能在實驗的初期就可以感覺出哪個選項中獎次數高，但無法一直持續的堅守在某個選項上。實驗四的設計是假設受測者採用的是相對次數的策略或類似的決策方法，那受測者在這個設計下會更容易選擇較好的選項，這裡實驗四在選擇 A 的比例上無法證明我們的假設。

表 4 選擇 a 比例 Wilcoxon 等級和檢定

選擇 a 比例	z	雙尾 p 值	n
Exp1 vs Exp2	-0.835941	0.403188	150
Exp1 vs Exp3	5.12347	< 0.001**	150
Exp1 vs Exp4	3.401	0.000671391**	150
Exp2 vs Exp3	6.65292	< 0.001**	150
Exp2 vs Exp4	4.05724	< 0.001**	150
Exp3 vs Exp4	-1.8822	0.059809*	150

4.3 轉換次數

我們計算受測者從兩個選項間轉換的次數 (A 選項轉換至 B 選項或 B 選項轉換至 A 選項都算是轉換一次)，轉換次數高除了代表著受測者容易認為上次所選的選項目不好，轉換次數高也代表受測者對採取探測 (explore) 行為的意願高。

圖 5 是四個實驗在全部 15 個 block 上的轉換次數平均，除了實驗三之外其他的實驗平均轉換次數趨勢都是向下的，實驗一、二與四平均轉換次數最小值都出現在最後一個 block，實驗二的平均轉換次數在 block8 以前都小於實驗一的平均轉換次數，實驗四、實驗二與實驗一比較不同資訊回饋的實驗中 Wilcoxon 等級和檢定平均轉換次數都是顯著的 p 值分別為 0.0171643 和 <0.001 。

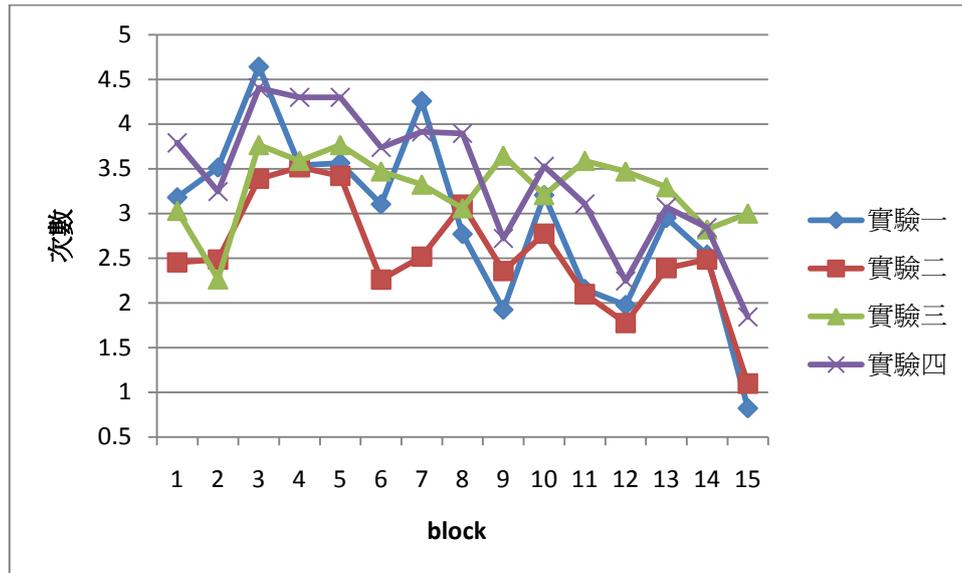


圖 5 轉換次數

實驗一整體的平均轉換次數為 2.94、實驗二整體的平均轉換次數為 2.54、實驗三整體的平均轉換次數為 3.29、實驗四整體的平均轉換次數為 3.40，在轉換的次數上不同資訊回饋的實驗差距都是顯著的，轉換次數最大是實驗四，轉換次數最小則是實驗一表示越多的資訊回饋下受測者轉換的次數越多，有較多的資訊時受測者會有更偏好搜尋的傾向。

表 5 轉換次數 Wilcoxon 等級和檢定

平均轉換數	Z	雙尾 p 值	n
Exp1 vs Exp2	-2.38317	0.0171643**	149
Exp1 vs Exp3	-8.28563	< 0.001**	149
Exp1 vs Exp4	-7.40094	< 0.001**	149
Exp2 vs Exp3	-6.36565	< 0.001**	149
Exp2 vs Exp4	-5.54953	< 0.001**	149
Exp3 vs Exp4	-0.535793	0.296051	149

註：樣本數為 149 原因為受測者的第一期選擇不算是轉換。

4.4 截斷點

截斷點策略是假設受測者是以“0”報酬的出現當作基準如果一個選項出現“0”的次數超過可以容忍的界限，下一次的選擇就會轉換到另一選項上，反之則下次的選擇會停留於目前的選項上。我們實驗截斷點 (cut point) 的計算是以每次受測者兩次轉換之為一個區間，計算這個區間內得到報酬“0”的次數來當作受測者的截斷點策略。

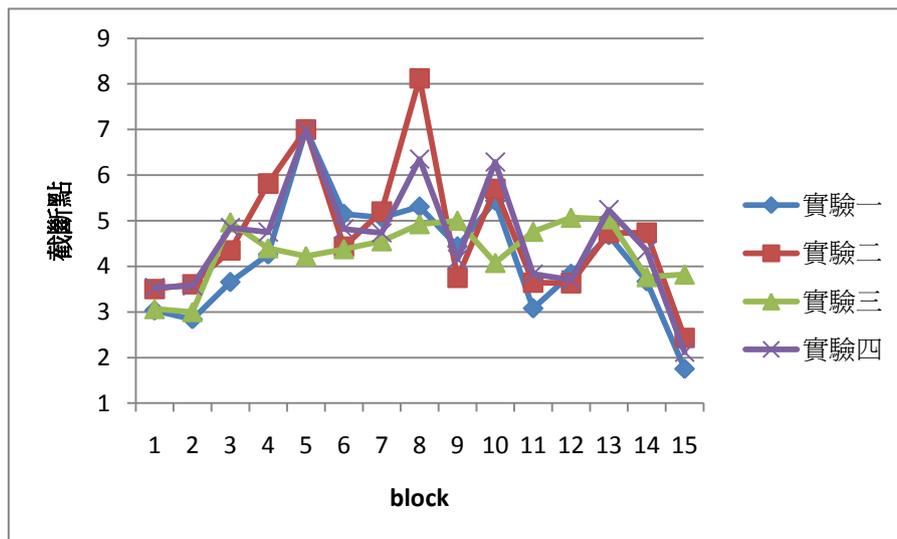


圖 6：四個實驗截斷點

四個實驗的截斷點最低的位置 block15 同時也是選擇 A 次數最高的時候，另外這種以計算報酬“0”出現次數為基礎的轉換行為四個實驗的趨勢也非常類似，似乎不會因為我們給受測者不同的資訊回饋而有不同。接著我們根據受測者把轉換的行為再細分成兩個類型：由 A 選項轉換至 B 選項、由 B 選項轉換至 A 選項。從實驗一、三與四裡皆看不出受測者是對 A 選項或是對 B 選項忍受“0”報酬出現次數的程度較高。從附錄中的圖 11 實驗一至四 a to b 截斷點與圖 12 實驗一至四 b to a 截斷點同樣的也看不出明顯的差異，表示個體不會因為不同資訊回饋就對 A 更有信心，對從 a 轉換到 b 採取更高門檻。

另外從實驗一、二與四各自的 a to b 和 b to a 截斷點也看不出一致的趨勢，但從圖 15 實驗三截斷點中可以看出受測者對不同選項可容忍報酬“0”出現的次數與其他的實驗不同，除了 block9 與 block10 之外實驗三的受測者對 A 選項可容忍報酬“0”所出現的次數一致的高於選項 B。

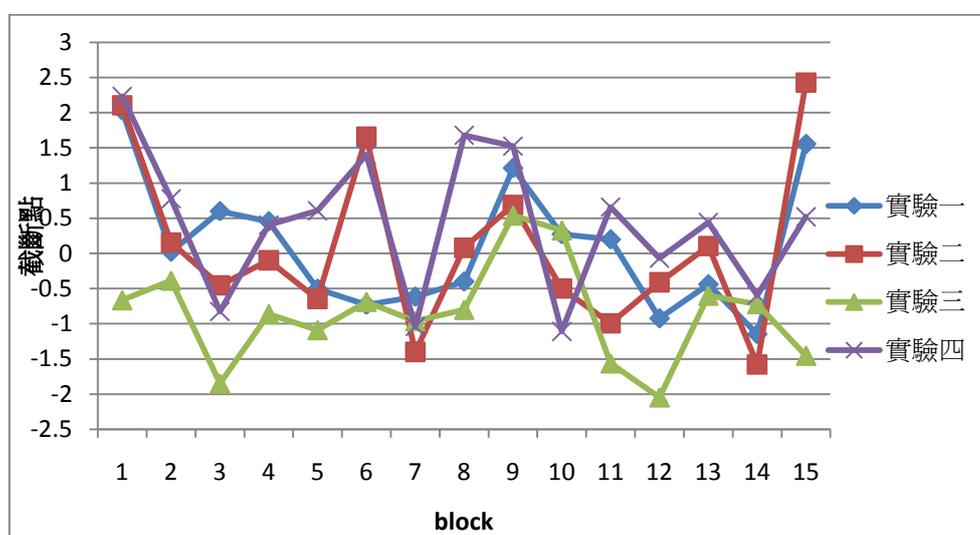


圖 7 四個實驗的 atob - btoa 兩個數列相減的趨勢

我們把 B to A 數列減去 A to B 數列的時間趨勢繪成圖 7，圖 7 中數線 0 以上的區域代表從 B 轉換到 A 行為出現前受測者所容忍報酬“0”出現次數大過從 B 轉換到 A 時受測者所容忍報酬“0”出現的次數，代表在數線高於 0 的 block 上受測者是對選項 B 信心較高；數線低於 0 以下的選項受測者

則是對選項 A 信心較高，簡單來說正值代表較喜歡 B，負值代表較喜歡 A。表 6 是用各實驗的 b to a 截斷點減去 a to b 截斷點的 t 檢定，同樣的如 t 值為正代表較偏好選項 B，負值代表較喜歡 A。

幾個結果顯著的有第二組實驗一與實驗三 ($p=0.003658$)，第四組實驗二與實驗三 ($p=0.01496$)，第六組實驗三與實驗四 ($p=0.0003794$)，雖然有最多資訊回饋的實驗四與實驗一和二相比是較偏好選項 A 的，但這樣的差異不明顯。顯著結果裡都是跟實驗三的平均數不同，而實驗三的設計主要的不同是在報酬函數上，顯示影響受測者對選項的信心不是資訊回饋而是報酬函數。

表 6 各實驗的 b to a 的截斷點減去 a to b 截斷點 t 檢定

組別	t (28) 值	p-value
Exp1 vs Exp2	0.0751089	0.9407
Exp1 vs Exp3	3.17161	0.003658*
Exp1 vs Exp4	-0.937332	0.3566
Exp2 vs Exp3	2.5929	0.01496*
Exp2 vs Exp4	-0.897222	0.3773
Exp3 vs Exp4	-4.0381	0.0003794*

從表 6 的幾個結果我們很難從截斷點裡面找到不同的反饋對個體在不確定性下選擇行為的影響，個體對某個選項的信心只受到不同報酬函數的影響，我們的受測者表現也不同於截斷點或 hot hand 模型所認為的對某個選項的信心是一個固定值，造成這樣的原因可能是受測者並非只使用同一種策略，他們在初期可能是較傾向探索到了後期轉為固定在某個選項上做開發的行為。另一種可能造成原因是個體間轉換次數的差異過大，導致我們無法在特定的指標上找出反饋的影響。

4.5 轉換次數分組

以上的幾個結果找不出不同資訊回饋的差異，這說明了個體進行不確定選擇時的決策方式不考慮我們實驗裡所給的資訊。另外，低轉換數或高轉換數有可能是一種個體在做不確定選擇時所採取的策略。爲了要驗證這個假設我們另外把四個實驗的資料共同放在一起處理，不以我們所給的資訊不同或報酬函數不同而是以四個實驗裡所有受測者的中位數分類，這裡我們分成高轉換組別與低轉換組兩個組別。

4.5.1 平均表現

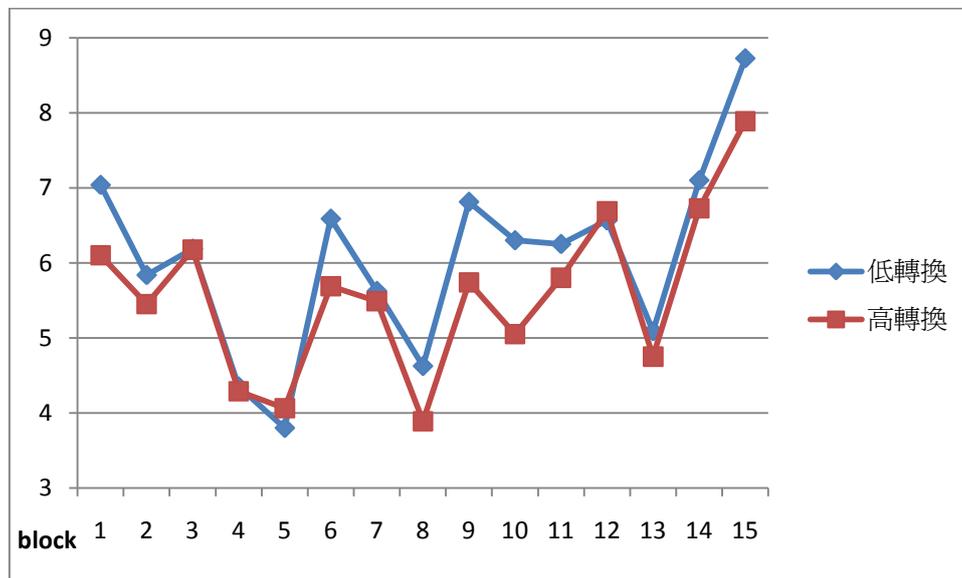


圖 8 轉換數分組的報酬

經過轉換次數分組後，低轉換組表現優於高轉換組但受測者報酬沒達到很大差異，差異最大是在 block10 上平均報酬差 1，整體的平均報酬高轉換組爲 83.78，低轉換組爲 90.75。再者兩者的平均報酬有相同之趨勢，平均報酬最高點都是出現在最後一個 block 上，表示說不管是高轉換受測者或是低轉換的受測者皆是在一開始的時候都是先探索，等到後期時才固定在一個選項上開發。我

們在報酬的結果上沒有看到很大的差異，是否是因為兩種不同轉換數組別間選擇狀況差別不大所造成。

表 7 選擇 a 次數等級和檢定

選擇 a 次數	Z	雙尾 p 值	n
高轉換組 vs 低轉換組	-7.85093	< 0.001**	150

4.5.2 選擇的狀況

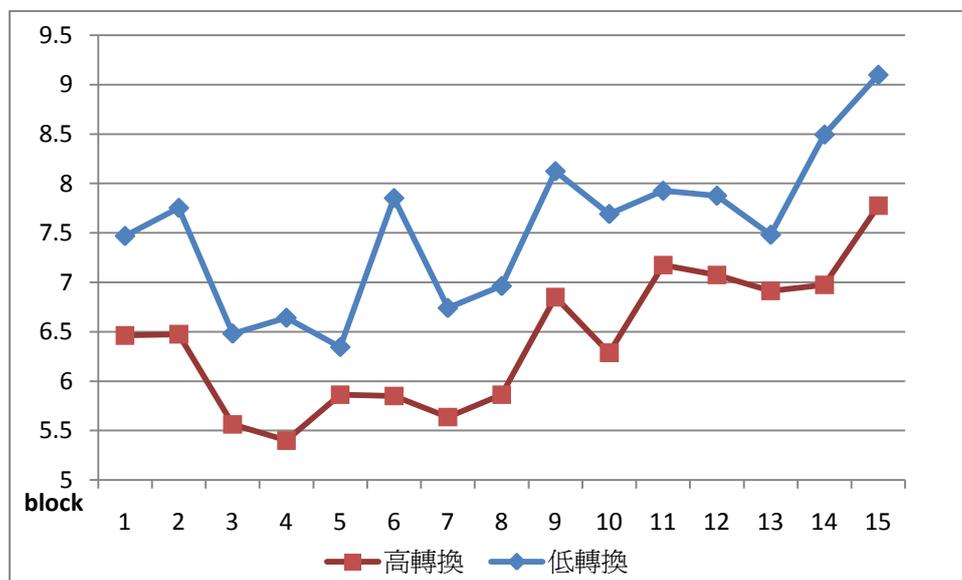


圖 9 高轉換組與低轉換組選擇 A 的平均次數

雖然前面低轉換數組別報酬與高轉換組沒有很大差異但從圖 9 我們可以看出，低轉換組選擇 A 的次數多於高轉換數組，不管是低轉換數或是高轉換數行為的受測者每 block 中選擇 A 的平均次數都超過所有選擇次數的一半，表示他們較偏好選項 A。

他們對選擇 A 的狀況是呈現一個上下波動的情況。從圖 9 來看低轉換組別的受測者選擇 A 的平均次數相對高於高轉換數的組別。在 block 3、4、5 和 10 上是受測者選擇 A 的平均報酬相對低的點因此高轉換數列與低轉換數列同時出現平均選 A 次數較低是合理的結果。

表 8 平均報酬 Wilcoxon 等級和檢定

平均報酬	Z	雙尾 p 值	n
高轉換組 vs 低轉換組	-2.03461	0.0418899**	150

下面我們更進一步的針對「過去 10 次所累積平均 A 中獎次數」對「選擇 A 次數佔所有選擇的比例」的影響，作一簡單的迴歸分析。結果如下表 9 和表 1

表 10 所示，得知高轉換數組過去 10 次累積平均 A 中獎次數 (H Acc.A) 對高轉換組選擇 A 次數佔所有選擇的比例 (HA) 影響是正向並且顯著的，另外在低轉換數組裡過去 10 次累積平均 A 中獎次數 (L Acc.A) 對低轉換組選擇 A 次數佔所有選擇的比例 (LA) 影響也是正向且顯著的，過去選 A 中獎次數有關係。

表 9 高轉換組選擇 A 比例和過去 10 選項 A 累積中獎次數的簡單回歸模型

$HA = \alpha_0 + \alpha_1 H \text{ Acc. A}$		
Regressors	coefficient	p-value
Constant	0.507023	<0.001
高轉換選擇 10 次平均 A 中獎次數 (H Acc. A)	0.00642379	<0.001
R-Square=0.149751, Adj. R-Square=0.143967, MSE=0.143967		

表 10 低轉換組選擇 A 比例和過去 10 選項 A 累積中獎次數的簡單回歸模型

$LA = \alpha_0 + \alpha_1 L \text{ Acc. A}$		
Regressors	coefficient	p-value
Constant	0.595198	<0.001
低轉換選擇 10 次平均 A 中獎次數 (L Acc. A)	0.0393324	<0.001
R-Square=0.203210, Adj. R-Square=0.197790, MSE=0.100874		

我們依照上面的方法分別對過去 9 到 1 次嘗試中選擇 A 且 A 中獎的次數，依照附錄表 12 與表 13 我們可以看到在高轉換組別，對選擇 A 比例有最高 Adj. R-squared 及最小 Akaike criterion 的變數是過去 2 次中 A 中獎次數；低轉換數組別則是過去 1 次 A 中獎次數。

從以上的分析可以知道個體的當次是否會選擇開發某個選項和過去這個選項的結果有關係，並且和短期的關係較大（過去 2 次和過去 1 次）。但由於圖 8 報酬的差別並不大，有可能不是因為報酬造成受測者會有不同的轉換數，而是受測者本身對探索行為的偏好不同，有人天生喜歡多探索造成較高轉換數、有人喜歡少探索造成低轉換數。接著我們進行截斷點的分析來解釋造成的原因是受測者得到報酬不同或受測者本身的差別。

4.5.3 截斷點

從圖 10 a to b 的截斷點的行為來看可以知道高轉換者與低轉換者在 block4 以前對兩個選項的信心沒有差別，但在 block4 之後低轉換數組別截斷值都高過高轉換數組別，兩數列的 t 檢定顯著 p 值為 0.0302 表示低轉換數組別對選項 A 的信心較高，同時低轉換的組別變動的幅度是高於高轉換組別。

如果決策者行為是屬於前者因為報酬高而有開發選項 A 行為，因為選擇 A 贏的機率較高受測者對選項 A 的信心也應該是遞增的，在圖 10 中我們看不到遞增的截斷點的趨勢，而是上下波動的情形表示有雖然個體已經可以知道選項 A 的報酬較大但仍不會放棄探索，因此會有這種代表信心的門檻值波動的情況。從這裡我們可以結論出決策時個體間採取探索行為偏好不同佔重要的影響，會造成轉換數上不同的原因只是因為個體對探索行為偏好不同。

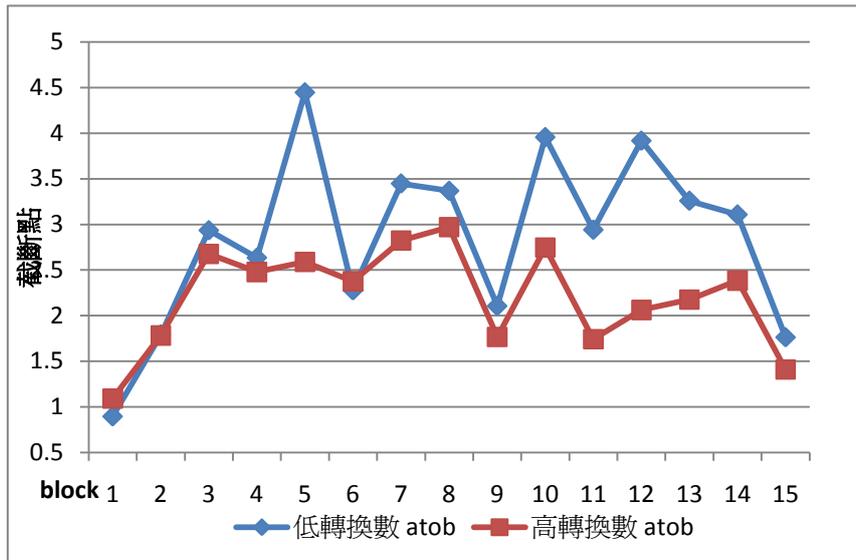


圖 10 高轉換組與低轉換組對從 A 選項轉換到 B 選項的截斷點行為

表 11 高轉換組與低轉換組截斷點 t 檢定

組別	t (28) 值	p-value
高轉換組 vs 低轉換組	-2.28339	0.0302

五、結論與展望

本文以 two-armed bandit 問題探討個體在面對不確定的選項時個體決策與學習狀況，two-armed bandit 問題是受測者面對兩個不確定選項，受測者對這兩選項的資訊只有透過實際採行才能從結果中獲得訊息。我們試著分析個體在面對這樣的決策過程，並從決策過程中，找尋面對不確定問題時，決策過程中受測者認知與所接收到的反饋所扮演的角色，透過實驗的方式探討不同反饋的下，選擇行為的狀況，我們用平均報酬、平均轉換次數、截斷點策略這幾個指標來衡量決策的行為。我們得到的幾個一般性的結果：

1. 不同資訊回饋的平均報酬，在我們的實驗裡看不出差異，顯示更多的資訊無法幫助受測者選擇有較高報酬的選項致使平均報酬沒有差異。
2. 轉換次數在受測者之間的差異很大但只有實驗二、三和實驗三、四的平均轉換次數不同。表示資訊回饋不同對受測者不確定行為上的轉換行為上影響不大。在實驗裡轉換次數的趨勢是會隨著嘗試 block 數增加而減少。
3. 最後我們檢視了兩個實驗的截斷點策略，四個實驗裡的受測者在這個表現上沒有一致的趨勢。
4. 在把受測者分成了高轉換組與低轉換組之後，截斷點的趨勢在 block3 之前是相同的，同時低轉換組受測者在截點變動的程度較大。

本篇目前的研究的結果是，不同的反饋對受測者的影響從前面的圖表看來是沒有影響的，這或許是因為受測者決策過程中忽略我們所給的資訊，所以在幾個表現的結果沒有差別。

根據我們研究的發現及研究時所碰到的問題後續的研究可能的方向是，針對每個受測者個人的特性分類，個人在解決不確定問題時可能因為某些個人特質(例:風險態度、工作記憶) 決策過程可能也會有所差異。另外由於本實驗的結果

兩種不同的反饋型式差異並不顯著，有可能是受測者的決策過程中並沒有我們所給對照組的變數，接下來應該考慮其他有可能是受測者決策過程中更具體的回饋型式。最後除了轉換次數與截斷值這兩行為指標在我們的實驗裡應該採用其他更可以解釋受測者和學習之間關係的指標進行分析。

參考文獻

1. Banks, J., Olson, M. and Porter, D. (1997), "An experimental analysis of the bandit problem," *Economic Theory*, 10, pp 55-77.
2. Becker, G .m., DeGroot, M.H., Marschak, J. (1986), "Measuring utility by a single-response sequential method," *Behavioral Science* , 9, 226-232.
3. Berg, J. E., Daley, L. A., Dickhant, J. W., O'Brain, J. R. (1986), "Controlling Preferences for Lotteries on units of experimental exchange," *Quarterly Journal of Economic*, 101, 281-306.
4. Bush, R.R., Mosteller, F. (1955), *Stochastic Models for Learning*. New York: Wiley.
5. Camerer, C. (1987), "Do biases in probability judgment matter in markets? Experimental evidence," *American Economic Review*, 77, 987-71.
6. Camerer, Colin F. and Robin M. Hogarth. (1999), "The effects of financial incentives in experiments: A review and capital-labor production framework," *Journal of Risk and Uncertainty*, 19, pp. 7-42
7. Einhorn, H. and Hogarth, R. (1985), "Ambiguity and Uncertainty in Probabilistic Inference," *Psychological Review*, 93, 433-461.
8. Fernie, G. and Tunney, R. J. (2006), "Some decks are better than others: The effect of reinforcer type and task instructions on learning in the Iowa Gambling Task," *Brain and Cognitor*, 60, 60, pp. 94-102.
9. Fischbacher, U. (2006), "Z-tree: Zurich toolbox for readymade economic experiments," *Experimental Economics*, pp. 171-178.

10. Gabiax, X., Laibson, D., Moloche, G. and Weinberg, S. (2006), "Costly information Acquisition: Experimental Analysis of a Boundedly Rational Model," *The American Economic Review*, 96(4), pp. 1043-1068.
11. Gans N., Croson R. and Knox G. (2007), "Simple Models of Discrete Choice and Their Performance in Bandit Experiments," *Manufacturing Service Operations Management*, 9, 383-408.
12. Grether, D. M. (1992), "Testing Bayes rule and the representativeness heuristic : Some experimental evidence," *Journal of Economic Behavior and Organization*, 17, 31-57.
13. Herrnstein, R. J. and Loewenstein, G.F. (1993), "G. F. Utility Maximization and Melioration: internalities in Individual Choice," *Journal of Behavioral Decision Making*, 6, pp. 149-185.
14. Gittins, J. C. and Jones, D. M. (1974), "A Dynamic Allocation Index for the Sequential Design of Experiments," *Progress in Statistics*, pp. 241-266.
15. J.J. Goodnow (1955), "Determinants of Choice-Distribution in Two-Choice Situations," *The American Journal of Psychology*, Vol. 68, No. 1, pp. 106-116.
16. Meyer, R. J. and Shi, Y. (1995), "Sequential Choice Under Ambiguity Intuitive Solutions to the Armed-bandit problem," *Management Science*, 41 (5) ,pp. 817-834.
17. Read, G., Loewenstein, G. and Kalyanaraman, S. (1999), "Mixing Virtue and Vice: Combining the Immediacy Effect and the Diversification Heuristic," *Journal of Behavioral Decision Making*, 12, 12, pp. 257-273.
18. Tunney, R. J. and Shanks, D. R. (2002), "A Re-examination of Melioration and Rational Choice," *Journal of Behavioral Decision Making*, 15, pp. 291-311.

19. Tunney, R.J. (2006), "Short article Preference reversals are diminished when gambles are presented as relative frequencies," *The Quarterly Journal of Experimental Psychology*, 59(9), pp. 1516-1523.
20. Tversky, A., "Intransitivity of preference (1969)," *Psychological Review*, 76, 31-48.
21. Tversky, A., and Kahneman, D. (1974), "Judgment under uncertainty: Heuristics and bias," *Science*, 185, 1124-31.
22. Zauberan, G. (2003), "The Intertemporal Dynamics of Consumer Lock-In," *Journal of Consumer Research*, 30, pp. 405-419.

附錄一 實驗使用之指導語

以下是實驗所使用的指導語，在每個實驗開始前會唸給受測者聽，在受測者閱讀完三個畫面的指導語後才會開始實驗。

● 指導語畫面 (一)

歡迎參加本次實驗

你將要進行一個連續 150 次的選擇

你有 100 點實驗幣來進行這 150 次的選擇

每期你需要決定選 A 或選 B

每次選擇完後你可以馬上知道得到報酬 "0" 點實驗幣或是 "1" 實驗幣

按"OK"繼續下個畫面

● 指導語畫面 (二)

A 有可能是好的按鈕, 也可能是壞的按鈕

B 有可能是好的按鈕, 也可能是壞的按鈕

好的按鈕得到報酬"1"機率高, 得到報酬"0"機率低

壞的按鈕得到報酬"1"機率低, 得到報酬"0"機率高

這代表如果這 150 期都選好的按鈕, 得到報酬會高於 150 期都選壞的按鈕

按"OK"繼續下個畫面

● 指導語畫面 (三)

你不用 150 次都選 A, 或 150 次都選 B

每一期你都可以自由的選 A 或 B

按"OK"開始實驗

附錄二 截斷點結果

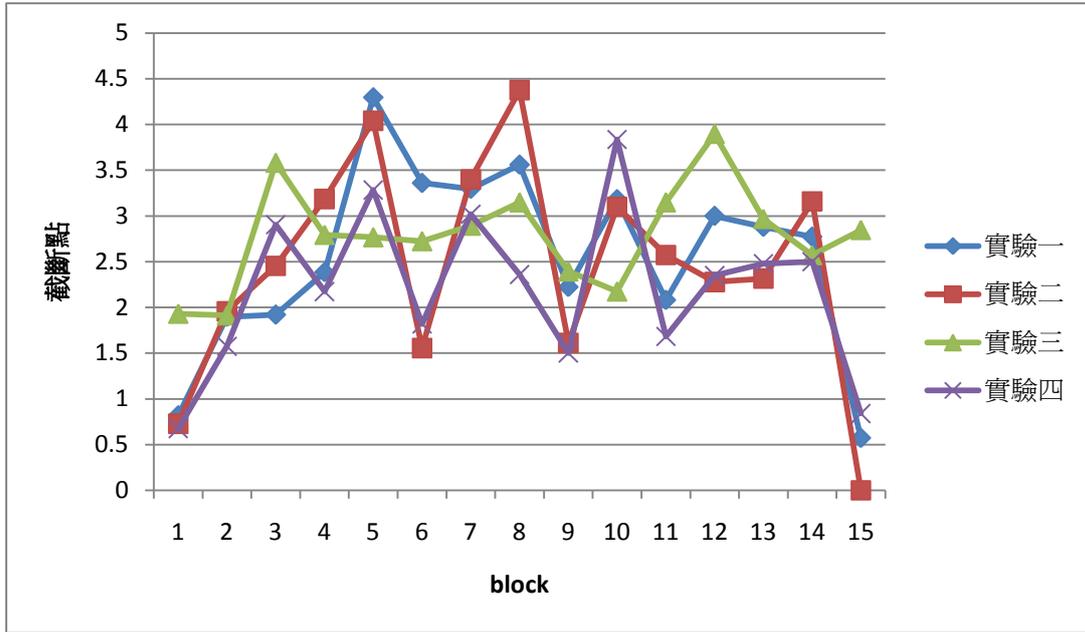


圖 11 實驗一至四 a to b 截斷點

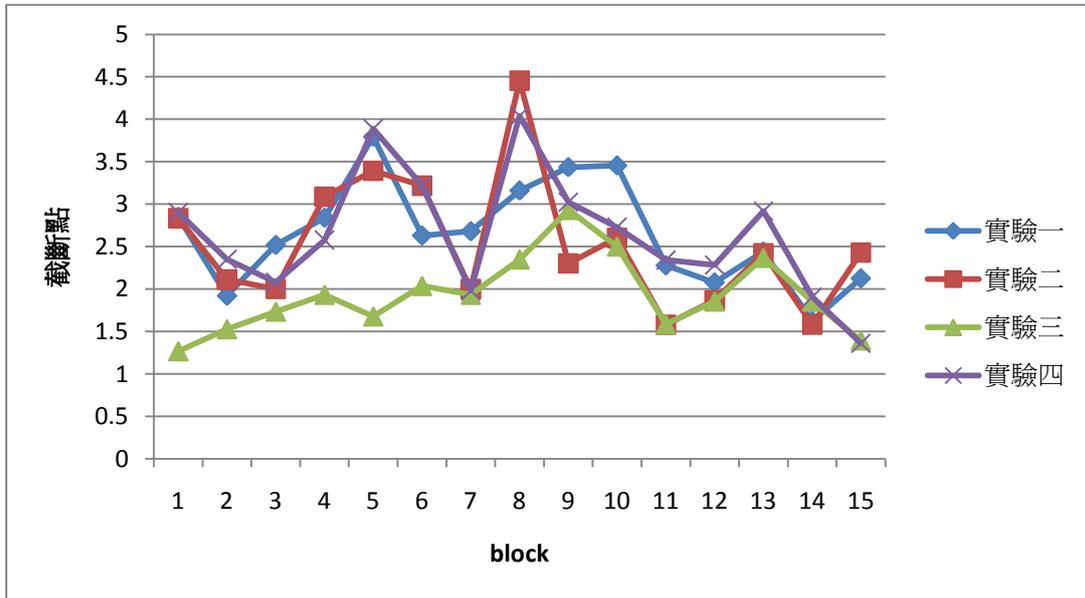


圖 12 實驗一至四 b to a 截斷點

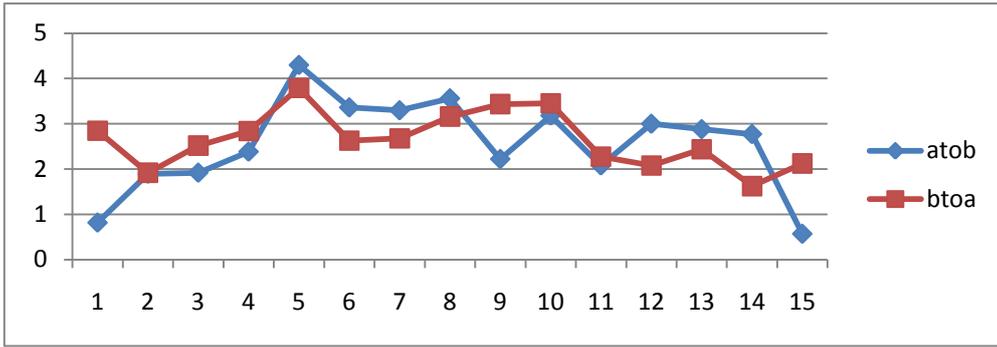


圖 13 實驗一截斷點

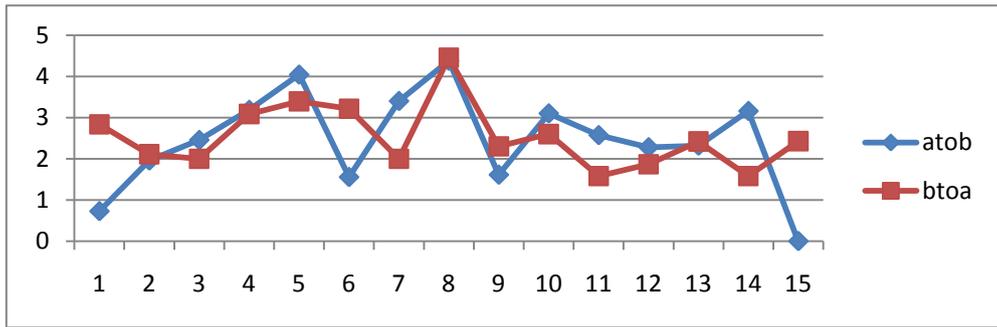


圖 14 實驗二截斷點

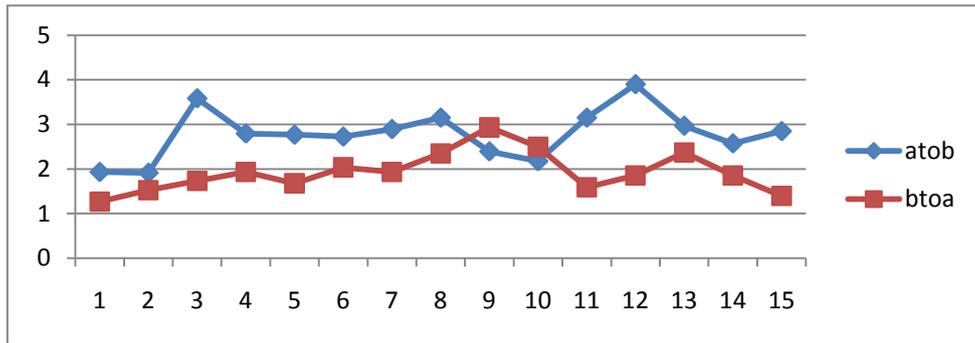


圖 15 實驗三截斷點

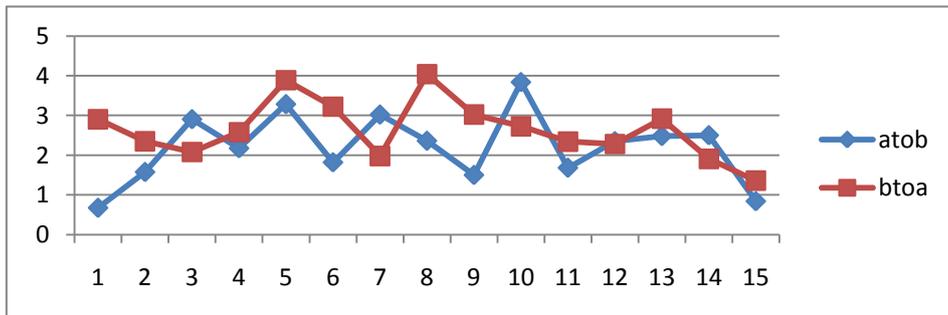


圖 16 實驗四截斷點

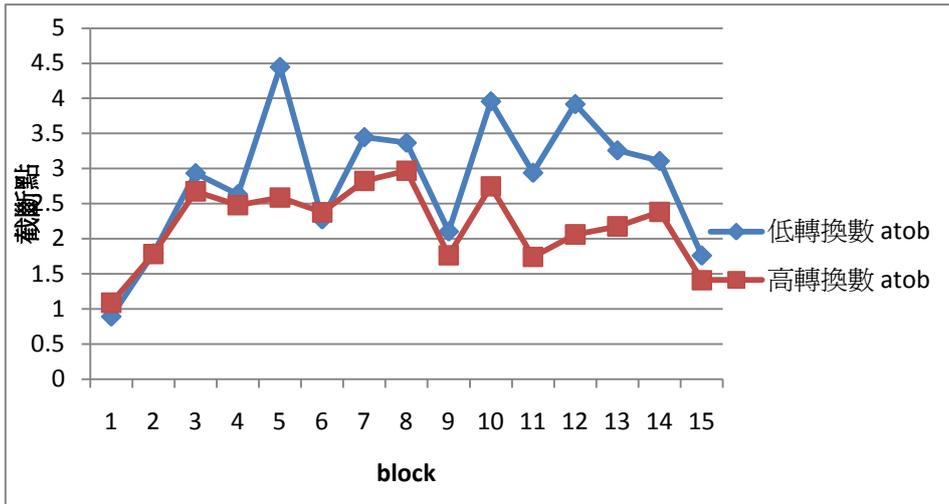


圖 17 轉換數分組後的 a to b 數列

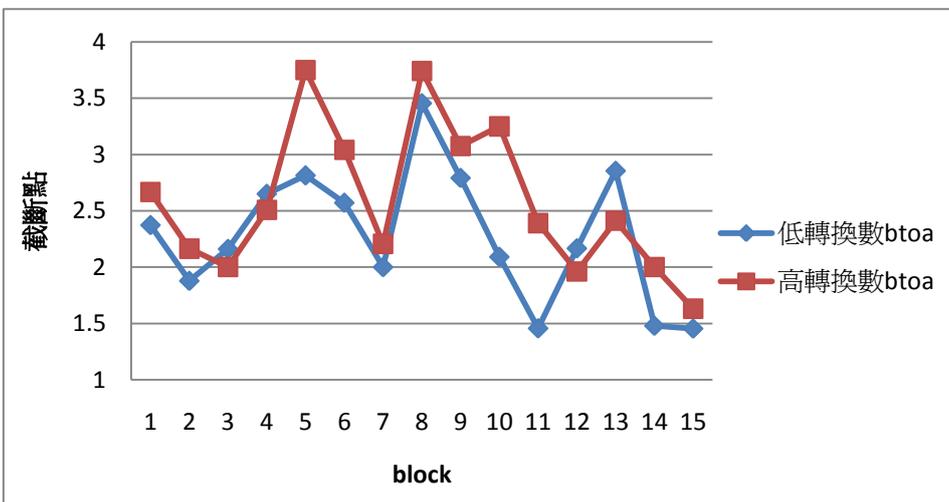


圖 18 轉換數分組後的 b to a 數列

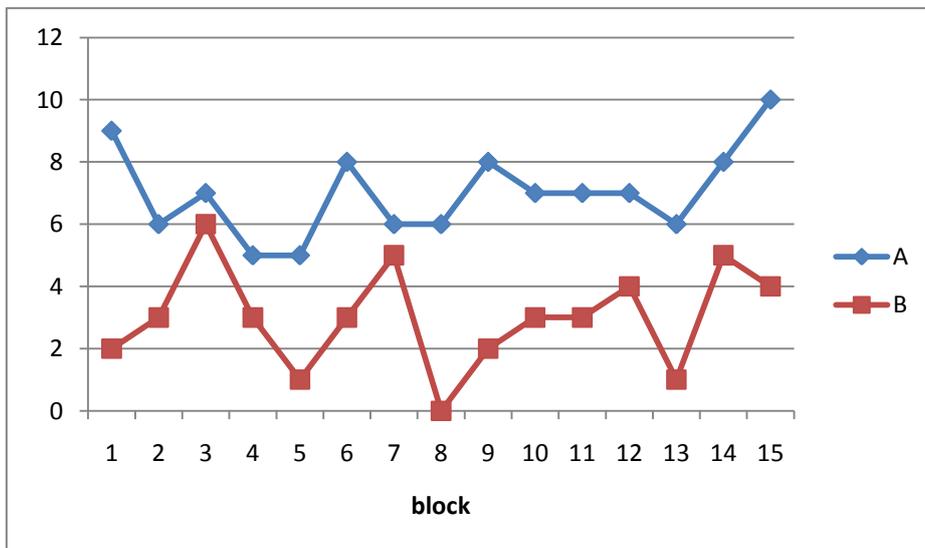


圖 19 A 與 B 實際的數列

附錄三 簡單回歸模型結果

表 12 高轉換組選擇 A 比例和過去 10 到 1 次嘗試選項 A 累積中獎次數的簡單回歸模型結果

解釋變數	係數	P-value	R-squared	Adj.R-squared	Akaike criterion
10	0.0274864 (0.00540190)	<0.001	0.149751	0.143967	-269.5999
9	0.0409450 (0.00628565)	<0.001	0.223999	0.218720	-283.2148
8	0.0460130 (0.00674807)	<0.001	0.242639	0.235121	-286.3760
7	0.0525174 (0.00715461)	<0.001	0.268223	0.263245	-291.9579
6	0.612586 (0.00789556)	<0.001	0.290527	0.285701	-296.5700
5	0.0729039 (0.00867090)	<0.001	0.324736	0.320142	-303.93333
4	0.0927706 (0.00959109)	<0.001	0.388923	0.384766	-318.8154
3	0.118170 (0.0114266)	<0.001	0.3421147	0.417209	-326.8875
2	0.171901 (0.0145766)	<0.001	0.486145	0.482649	-344.6344
1	0.291711 (0.0247559)	<0.001	0.485746	0.482248	-344.5190

表 13 低轉換組選擇 A 比例和過去 10 到 1 次嘗試選項 A 累積中獎次數的簡單回歸模型結果

解釋變數	係數	P-value	R-squared	Adj.R-squared	Akaike criterion
10	0.0393324 (0.00642379)	<0.001	0.203210	0.197790	-258.7467
9	0.0369953 (0.00563800)	<0.001	0.226547	0.221286	-263.1760
8	0.0412323 (0.00606905)	<0.001	0.238960	0.233782	-265.5856
7	0.0463046 (0.00653394)	<0.001	0.254649	0.249578	-268.6903
6	0.0534137 (0.00716624)	<0.001	0.274271	0.269334	-272.6654
5	0.0627109 (0.00790827)	<0.001	0.299605	0.294840	-277.9598
4	0.0796855 (0.00875163)	<0.001	0.360606	0.356256	-291.5371
3	0.104325 (0.0101650)	<0.001	0.3421147	0.417436	-305.4046
2	0.145945 (0.0129065)	<0.001	0.486145	0.482649	-318.1525
1	0.242300 (0.0208398)	<0.001	0.485746	0.482248	-322.0689

附錄四 轉換次數、選擇 A 次數與報酬的 Box-plot 圖

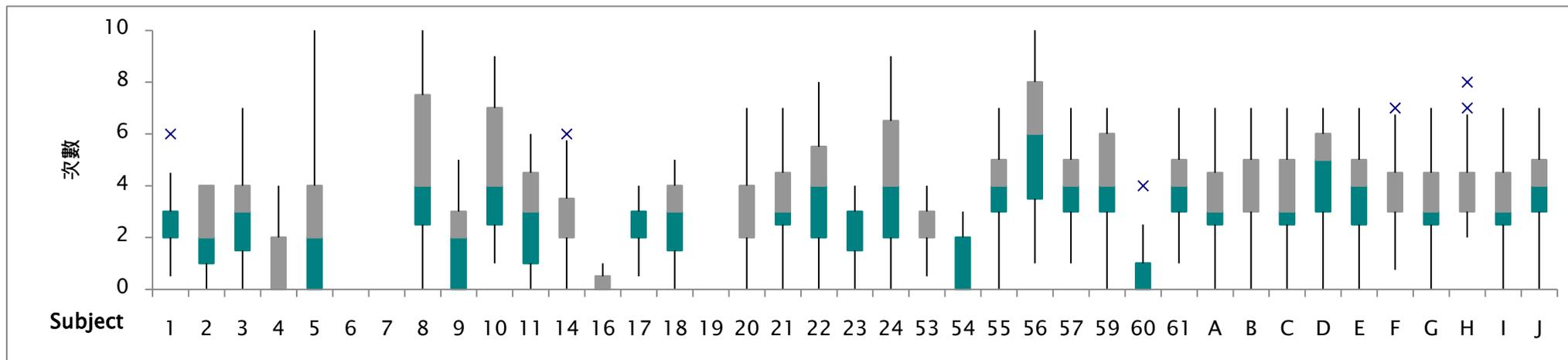


圖 20 實驗二轉換次數 Box-plot

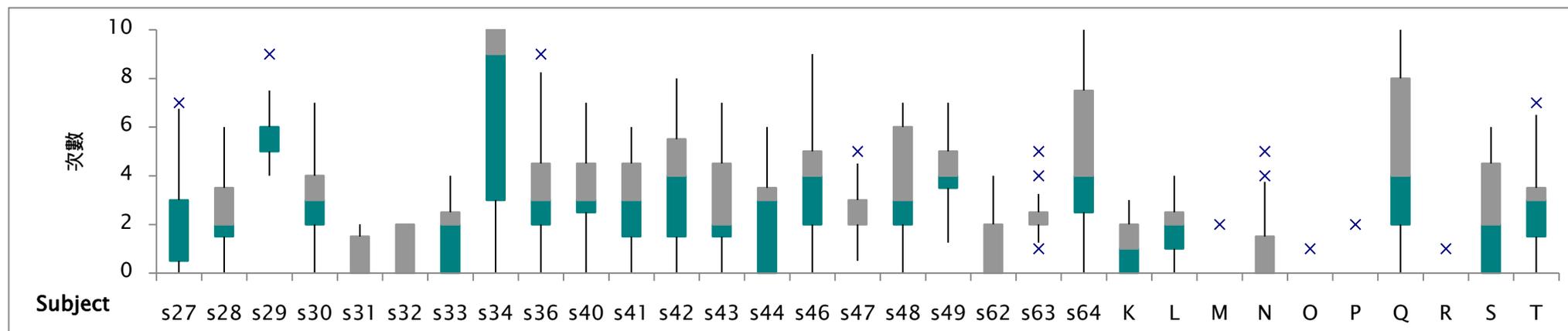


圖 21 實驗二轉換次數 Box-plot

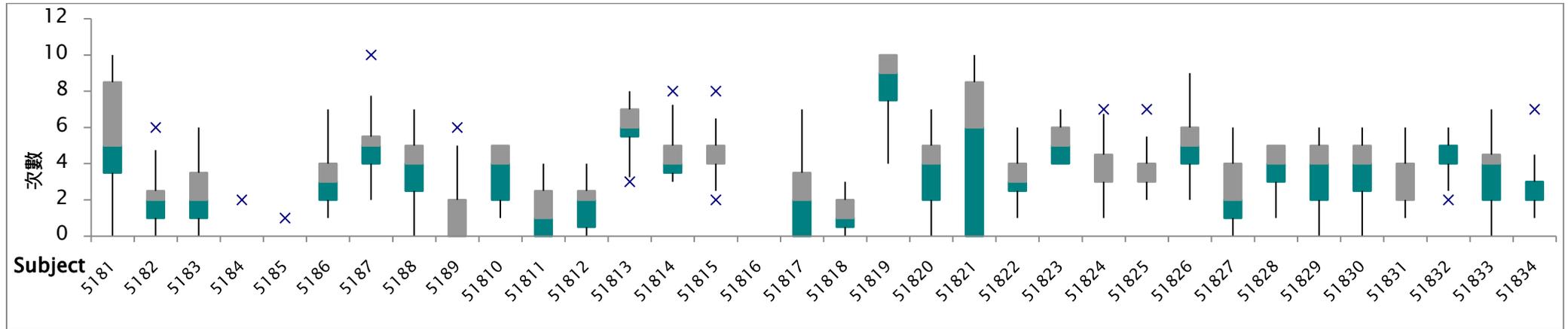


圖 22 實驗三轉換次數 Box-plot

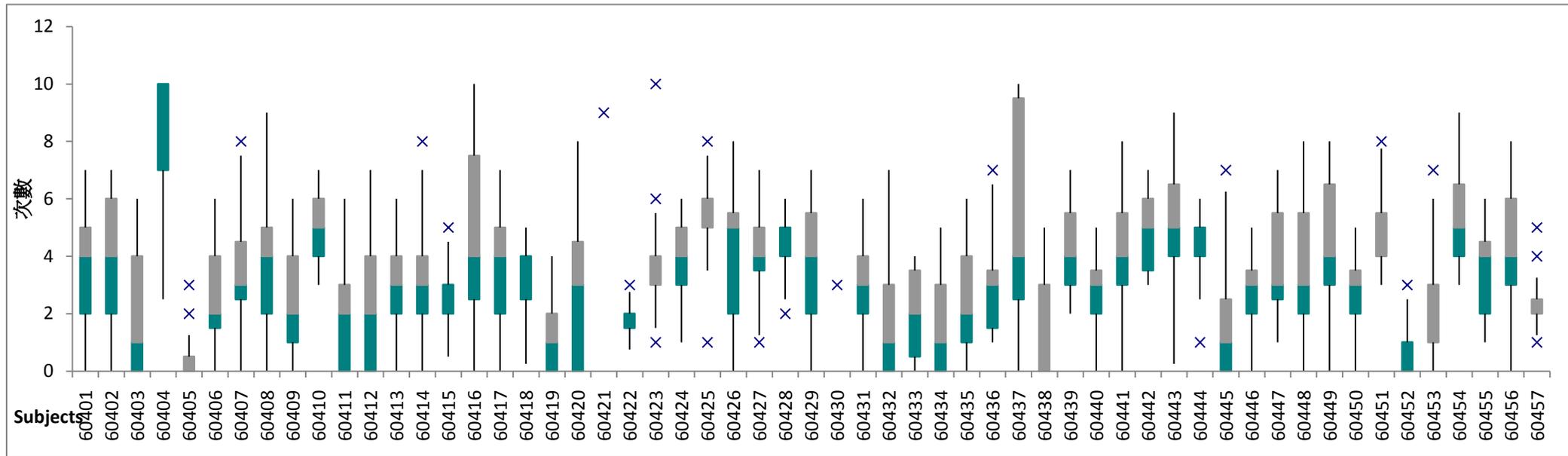


圖 23 實驗四轉換次數 Box-plot

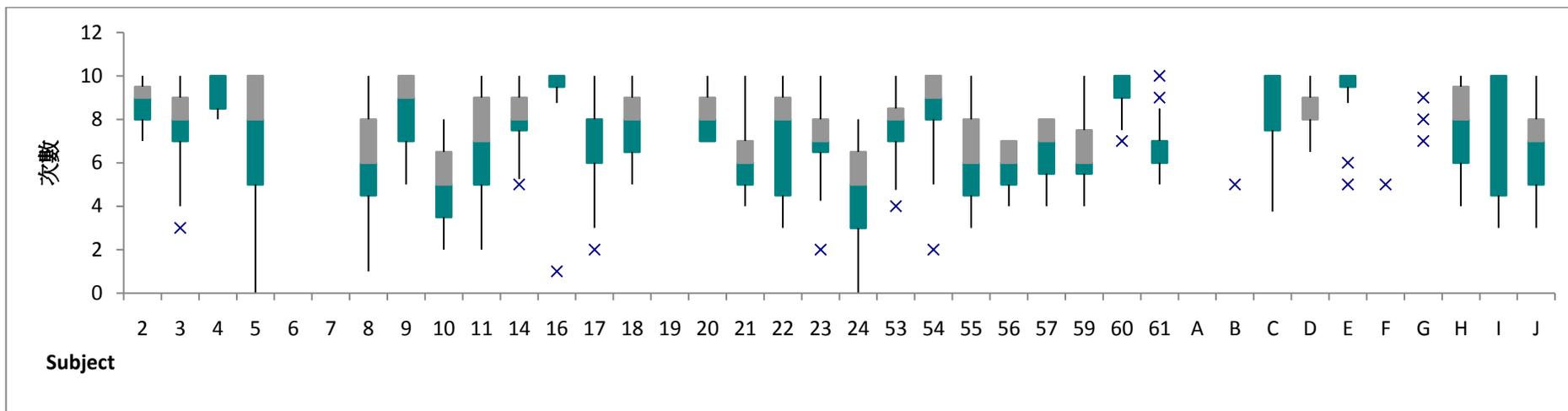


圖 24 實驗一選擇 A 次數 Box-plot

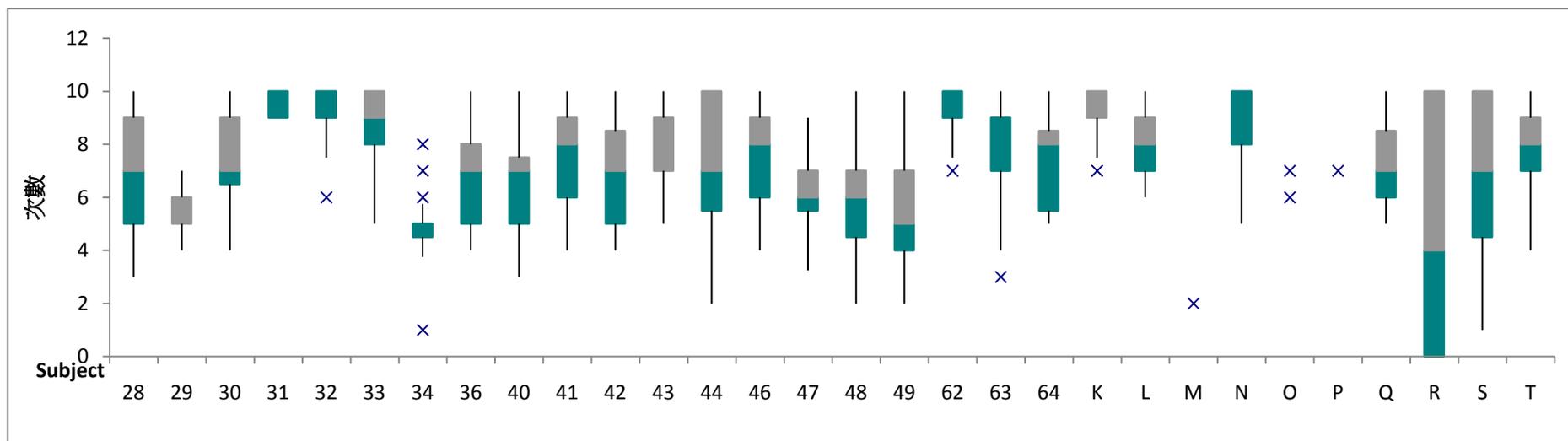


圖 25 實驗二選擇 A 次數 Box-plot

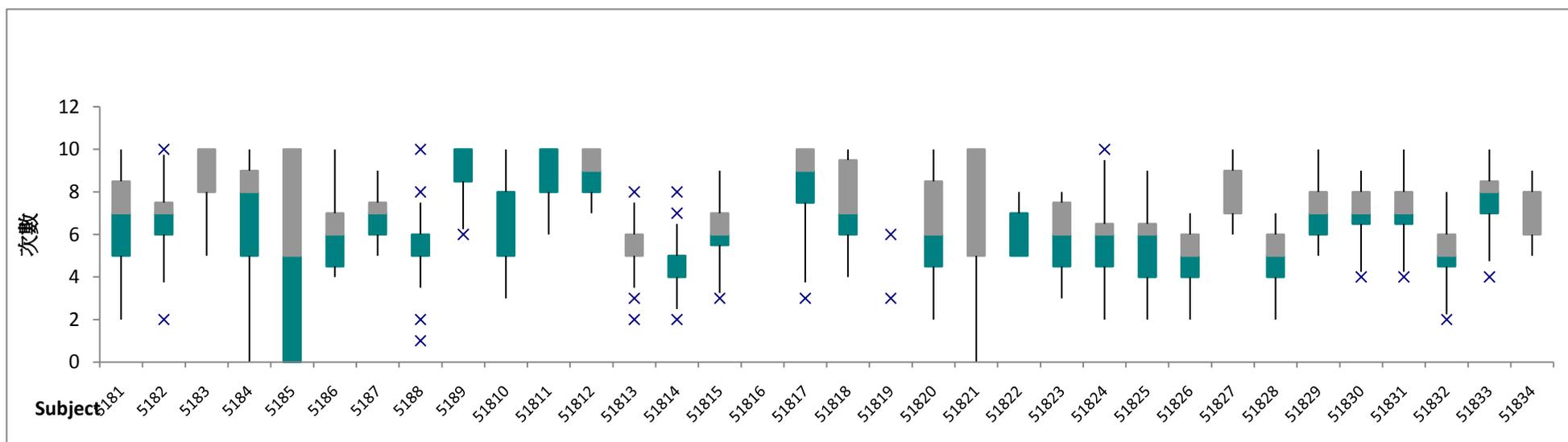


圖 26 實驗三選擇 A 次數 Box-plot

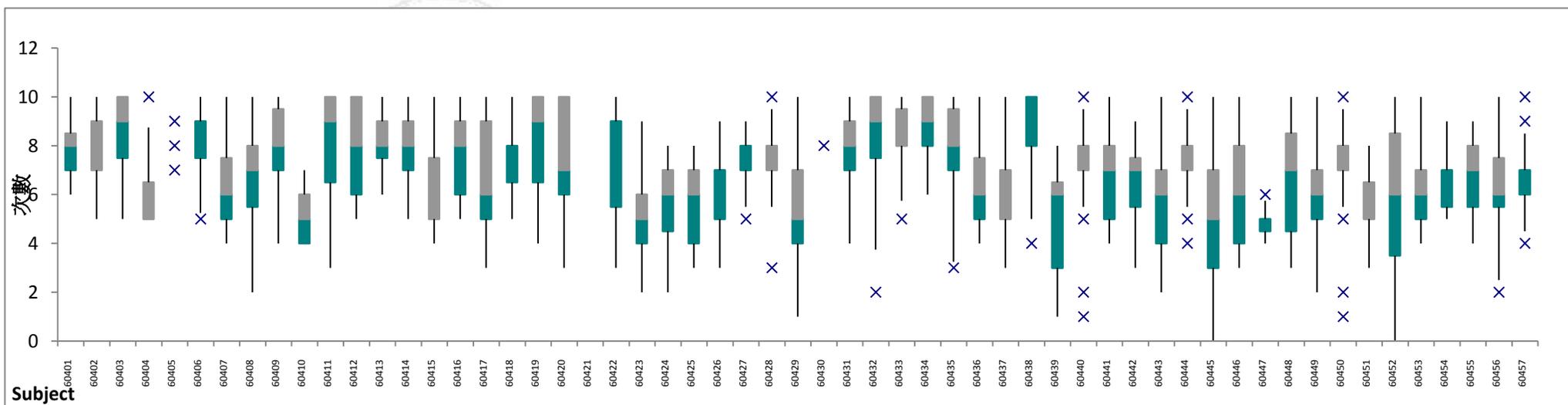


圖 27 實驗四選擇 A 次數 Box-plot

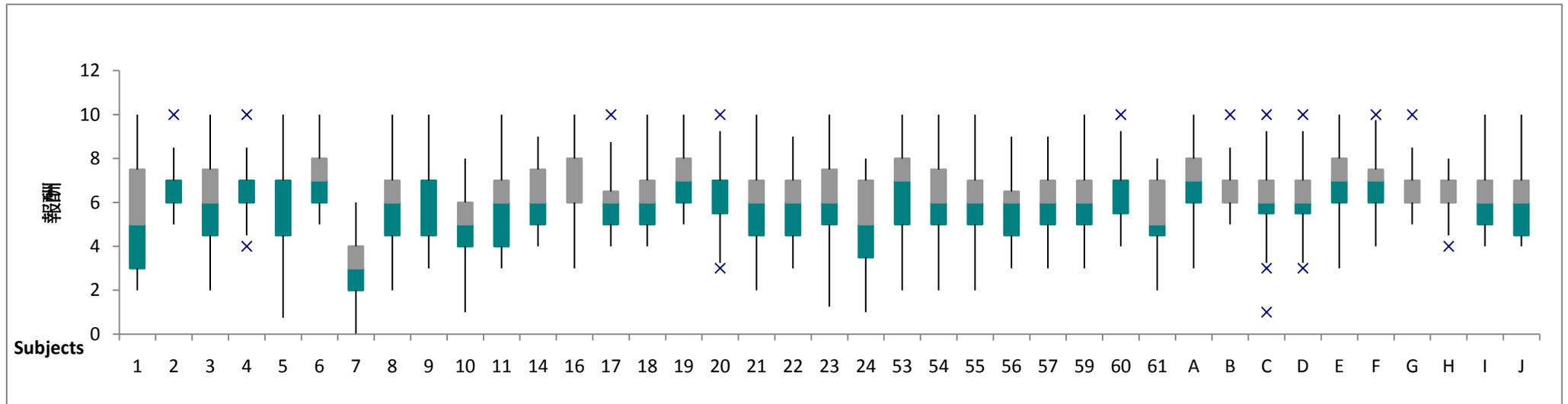


圖 28 實驗一報酬 Box-plot

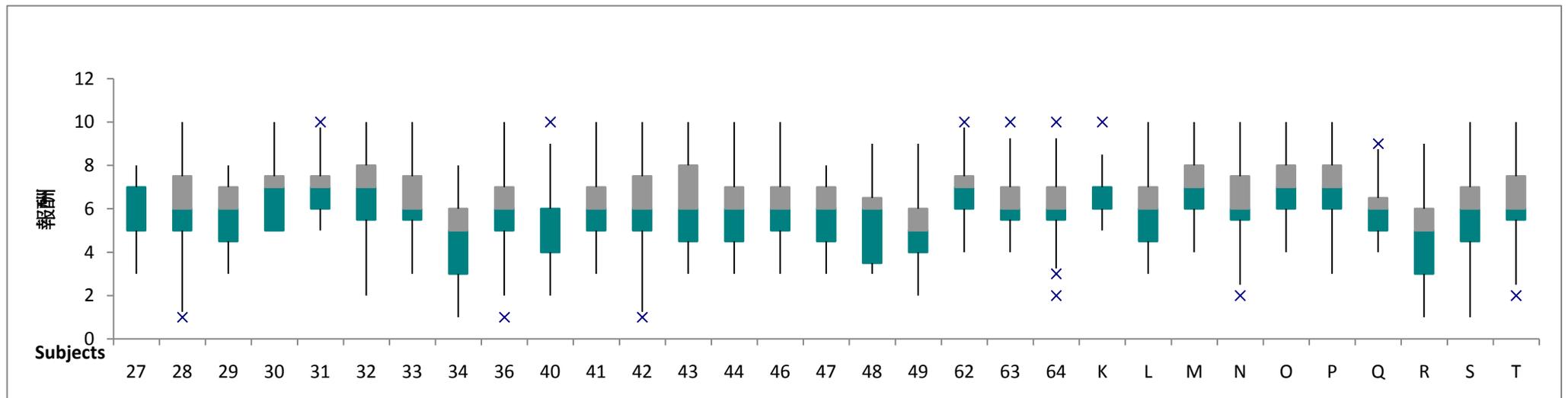


圖 29 實驗二報酬 Box-plot

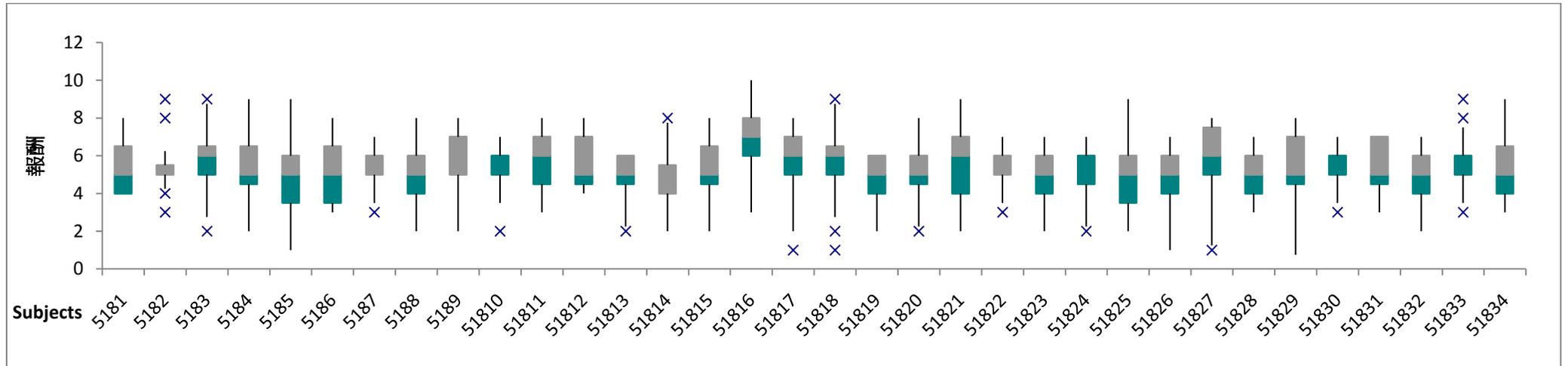


圖 30 實驗三報酬 Box-plot

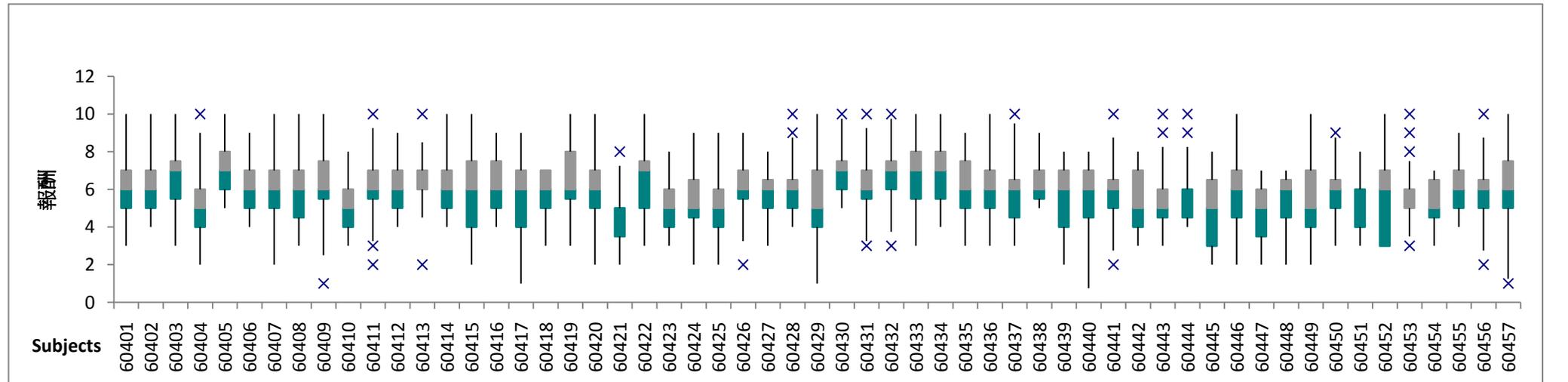


圖 31 實驗四報酬 Box-plot