

因果推論在政治學的發展與應用：以調查實驗為例

黃紀[†]

國立政治大學政治學系

摘要

「反事實因果推論模型」在社會科學研究中掀起了一場「可信度革命」，政治學自不例外。本文的目的，在回顧本世紀以來因果推論的黃金標準——實驗法——與傳統之調查研究結合而成之「調查實驗」，在政治學的快速發展與應用。政治學在實驗法的發展與應用起步雖然較慢，但 2000 年之後，實驗法開始躋身為政治學實證研究的主流之一，2010 之後更是呈快速成長，其中尤以調查實驗脫穎而出。本文聚焦於近十餘年來，政治學調查實驗中處理多面向因素的兩種實驗設計：短文實驗（vignette experiment）與聯合實驗（conjoint experiment），說明其研究設計之結構與應用。檢視近十年發展與應用最快速的聯合實驗，第一階段的可信度革命在確立因果推論的內在效度方面，已推陳出新。展望未來之發展與應用，勢必朝推廣因果關係之外在效度上力求突破。

關鍵詞：因果推論、實證政治研究、調查實驗、短文實驗、聯合實驗。

JEL classification: C90.

[†]通訊作者：黃紀
E-mail: chihuang@nccu.edu.tw

1. 導論

「反事實因果推論模型」(counterfactual model of causal inference) (Rubin 1974) 在社會科學的實證研究中掀起了一場革命，經濟學者 Angrist and Pischke (2010) 回顧 1990 年代以來計量經濟學在因果推論的信度上大幅改進，稱之為「可信度革命」(credibility revolution)，並將此進展歸功於嚴謹的研究設計，包含隨機分組實驗及準實驗設計(如工具變數、斷點迴歸、雙重差分等)。政治學自不例外於這場革命 (Ashworth, Berry and de Mesquita 2021；黃紀等 2007；黃紀 2008；黃紀 2010；黃紀 2013)，且其發展極為快速。本文的目的，在回顧本世紀以來因果推論的黃金標準——實驗法——與傳統之民意調查結合而成之「調查實驗」(survey experiments) (Mutz 2011)，在政治學的發展與應用。

實驗法具有兩大特徵：一為研究者能操作施測之「因」(cause or treatment)，或至少能操作受試者對「因」之感知(perception)；二為能依研究設計將受試者隨機分組。由於這兩個特徵最能確保「因」在統計上獨立於潛在後果(potential outcomes)，故常被推崇為因果推論之黃金標準 (Imai, King and Stuart 2008)。例如在二分類的因，平均而言，實驗組與控制組互為本尊與分身：控制組成為「若實驗組未暴露於因，其平均後果會如何」的分身，而實驗組就成為「若控制組也暴露於因，其平均後果會如何」的分身。因此兩組「果」的平均差值，就是平均因果效應(average treatment effect；ATE)。

相對於其他社會科學，政治學在實驗法的發展與應用起步較慢 (Kinder and Palfrey 1993)，但近年來則快速發展。Mize and Manago (2022) 計算政治學、經濟學、社會學的各兩份頂級國際期刊¹自 1990 年至 2020 年發表的論文中，採用實驗法的論文所佔的比例，據以比較實驗法在三個學門中的發展趨勢。圖 1 顯示，2000 年以來，實驗法在三個學門中都逐漸成長，而其中政治學的成長速度更是後來居上，從 1990 年的 2.2% 躍升至 2020 年的 18.9%，實驗法的佔比已不亞於經濟學與社會學。

實驗法開始躋身於政治學實證研究的主流之一，也展現在學術社群的組織上。2010 年美國政治學會之實驗研究群(experimental research section) 首度集會，並自 2014 年起正式發行「實驗政治學期刊」(*Journal of Experimental Political Science*)，

¹ Mize and Manago (2022) 收錄的三個學門期刊包括：

政治學：*American Political Science Review* (APSR), *American Journal of Political Science* (AJPS)
經濟學：*American Economic Review* (AER), *Econometrica* (Econ)

社會學：*American Sociological Review* (ASR), *American Journal of Sociology* (AJS)

圖 1 之圓圈、菱形、三角形為論文在三個學門中所佔之比例，線條則是 Mize and Manago (2022) 以半年為帶寬(bandwidth)進行局部加權迴歸(local weighted regression；lowess)之平滑時間趨勢線。

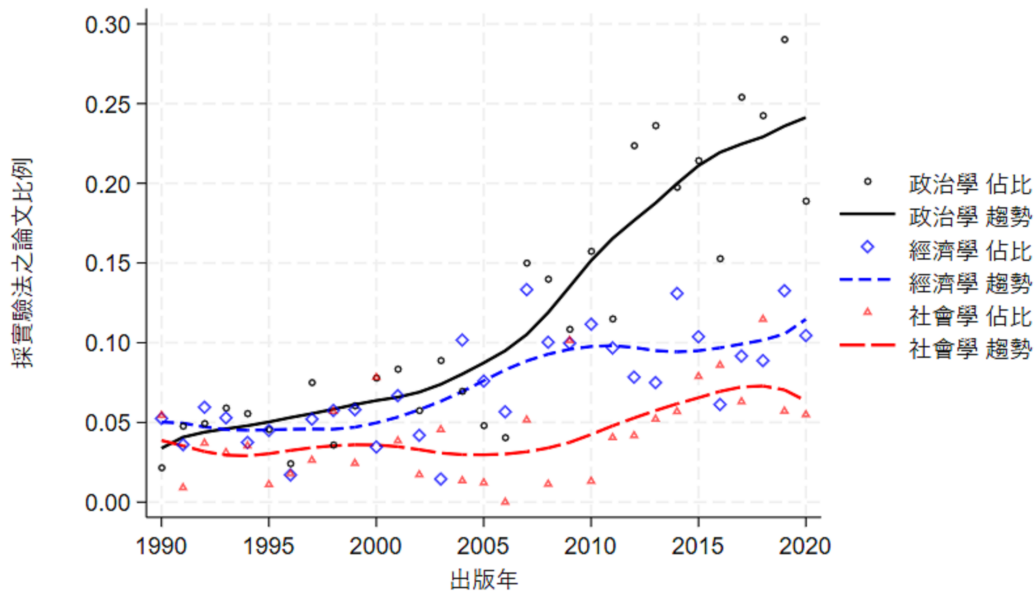


圖 1: 政治、經濟、社會學兩份頂級英文期刊中採實驗法之比例：1990-2020。

實驗法的社群力量儼然成型。

實驗法又可分成三大類：實驗室實驗 (laboratory experiments)、實地實驗 (field experiments)、調查實驗 (survey experiments) (參見 Morton and Williams 2010; Hyde 2015; Bol 2018; Sniderman 2018; Thomas 2024 等)。圖 2 進一步將政治學的兩份旗艦型英文期刊² *APSR* 及 *AJPS* 在 1990 至 2020 期間，採實驗法發表的論文進行分類，圖 2 顯示三種實驗設計的演變。在 2000 年前，政治學有限的實驗法論文之中，以實驗室實驗及調查實驗稍多，而實地實驗才剛萌芽 (Gerber and Green 2000; Gerber and Green 2012)。但 2000 年之後，實地實驗法開始躋身為政治學實驗法的主流之一。2010 年是個分水嶺，實驗室實驗比例漸下滑，而實地實驗與調查實驗則快速上升，其中調查實驗的成長幅度尤其明顯。

綜合 Druckman, Green, Kuklisky and Lupia (2006)、Druckman, Green, Kuklisky and Lupia (2011)、Druckman and Green (2021)、Druckman (2022) 及 Green (2022) 的回顧，可將 2000 至今區分為兩個發展階段：

1. 實驗政治學 1.0 (2000-2009)：調查實驗與實地實驗快速成長，追平了傳統的實驗室實驗。

²國內中文政治學期刊採實驗法的論文，近年也開始嶄露頭角。搜尋《台灣政治學刊》、《政治學報》、《選舉研究》三份期刊自 2011 至 2022 的論文，共有 2 篇採實驗法：包括調查實驗 1 篇 (楊光、鄭堉尹 2019)、實驗室實驗 1 篇 (郭銘傑 2022)。

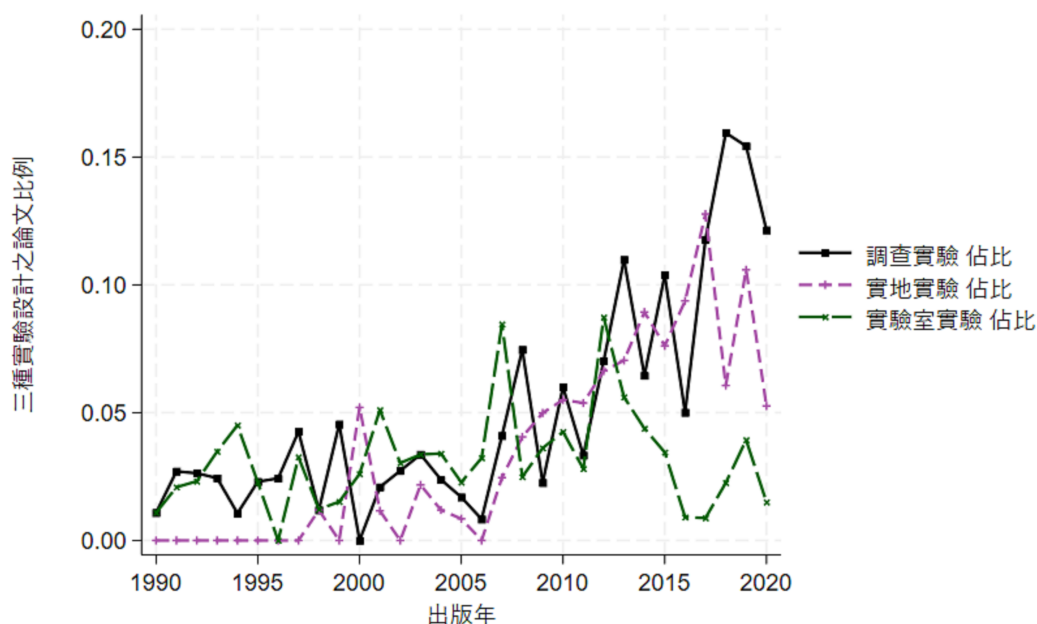


圖 2: 政治學 *APSR* 及 *AJPS* 採實驗法之論文中三種實驗設計之比例：1990-2020。

2. 實驗政治學 2.0 (2010-)：調查實驗與實地實驗繼續成長，但其中調查實驗更超越了實地實驗。

有鑑於調查實驗的快速成長，本文將探討政治學界近十餘年在這方面的發展與應用³，尤其聚焦於處理多面向（multidimensional）因果推論之短文實驗與聯合實驗設計。儘管特定的方法通常是在處理特定的議題脈絡時開發，但往往提出後隨即快速擴展至其他議題，故以下之論述，將以調查實驗方法的發展為主軸，先說明其方法與設計，再兼論其在各次領域中之應用。

2. 多因子調查實驗

一般的實驗設計，只針對一個變數（單因子）將受測者隨機分成兩組或數組，雖然簡單明瞭，但研究的題材侷限於較為單純的因果關係。若研究議題的「因」涉及多個面向，則宜採用能兼顧複合因果效應（composite treatment effects）的實驗設計。這種設計在不同學科中有其源遠流長的歷史，其中最早在行銷學發展出來的

³政治學常見之「列項實驗」（list experiments，亦稱 item count technique；ICT）雖也採隨機分組，但其研究焦點是以「間接問法」（indirect questions）處理敏感題，而非因果推論本身，故未列入本文之討論。有關列項實驗之回顧，請參見 Blair et al. (2020)、Thomas (2024)、黃紀 (2023)。國內列項實驗之應用，請參見 楊光、鄭棻尹 (2019)。

「聯合分析」(conjoint analysis) (Green and Rao 1971) 及計量經濟學中的「表述抉擇實驗」(stated choice experiments) (Louviere, Hensher and Swait 2000; Hensher, Rose and Greene 2015)、以及在社會學中的「多因子調查」(factorial surveys) (Rossi 1979; Wallander 2009; Atzmüller and Steiner 2010; Auspurg and Hinz 2015) 等。

2.1 短文實驗 (vignette experiment)

率先將多面向之隨機實驗題組鑲嵌到民意調查問卷中的，是社會學家 Rossi (1979)。該實驗題組是以短文 (vignette) 敘述某議題假想之人、物、情境，由研究者針對若干個屬性 (attributes)⁴ 的不同類別 (levels) 進行組合，形成幾種短文版本後，隨機分派給受訪者，詢問其對短文所述內容之評價，據以測量社會對該主題之判斷原則 (judgement principles)，並可進而搭配問卷中受訪者的協變數 (covariates) 資料，分析有無次群體差異 (subgroup differences)。為了確保不同組的受訪者接受的「因」只在屬性類別有差異，其餘皆相同，短文常刻意設計成類似「鏤空之模版」格式，模版基本描述文字均相同，僅在鏤空之處插入屬性隨機分派的類別 (參見表 1 之實例)。由於實驗題組是由短文構成，故常稱為「短文式實驗」(vignette experiments)；又因這類調查實驗涉及兩個或多個屬性，因此也稱為「多因子調查」。

短文式實驗不僅常見於社會學及心裡學文獻，政治學者也常採用。例如 Tomz (2007) 以短文式實驗檢測國際關係理論中的「觀眾成本」(audience costs) 說，亦即民主國家的領袖在處理外交軍事危機時，若先採威脅動武的嚇阻政策，事後卻食言退縮 (backdown)，將遭受國內民眾的責難。過往以一般民調所做的實證研究莫衷一是，難有定論。Tomz (2007) 另闢蹊徑，設計之實驗題組，假想甲國出兵佔領鄰近之乙國，短文中包含了軍事危機的 4 個脈絡屬性，每個屬性各有兩類，再搭配兩種不同結果，總共 5 個 2 分類的屬性，構成 $2^5 = 32$ 種組合的情境：

1. 甲國的政體 (2 類)：獨裁、民選；
2. 甲國出兵的動機 (2 類)：擴權掠奪資源、多年歷史宿怨；
3. 甲國的軍力 (2 類)：強、弱；
4. 對美國的利益 (2 類)：有害、無影響；
5. 結果 (2 類)：美國始終不介入、總統先宣稱會出兵擊退甲國但卻食言。

⁴在調查實驗之文獻中，「屬性」(attribute) 與「因子」(factor) 常視為同義字。

受訪者在閱讀隨機分派的一種組合情境後，以 1 至 7 的尺度給予總統評價，分數越高評價越正面。例如 32 種組合情境中的一種，其短文內容與評價題，參見表 1。

Tomz (2007) 的實驗，發現美國總統若先採威脅動武的嚇阻政策，但事後卻食言退縮，比起一開始就不介入該事件，還更受國內民眾的責難與負評。之所以有「觀眾成本」，主要是因為民眾在意美國或其領導人之國際聲望。而「觀眾成本」的考量，也牽動民主國家領導人因國內壓力對其國際安全承諾的信守 (commitment)。

短文式實驗在政治學中也常應用於較有爭議的內政議題研究，例如政治貪腐 (corruption) 對民眾選舉課責 (electoral accountability) 的影響 (Carreras and Vera 2018)、框架效應 (framing) 對右翼民粹之移民政策觀點的影響 (Kaufman 2019) 等 (見 Thomas 2024)。

短文式實驗的優點，在於以一段精簡的文字呈現假想的標的或情境，讓受訪者表達自己的偏好或評價。不過 Hainmueller, Hangartner, and Yamamoto (2015) 的研究指出，過去學界常採用的「單一短文設計」(single-vignette designs) 往往不能確保受訪者對該則短文的專注，因此測得的結果，準確度不如「兩個選項並呈的聯合實驗設計」(paired-conjoint designs)。此外，因每位受訪者只回應屬性類別隨機分派的一則短文，短文之屬性類別因人而異，故難以深入比較分析屬性之間的抵換關係 (trade-offs between attributes)。

表 1: 短文式調查實驗問卷中情境組合之一示例。

<p>某國出兵想佔領鄰國。該攻擊國的獨裁領袖因為想掠奪鄰國資源發動侵略。由於侵略國軍力強大，美國必須動用相當武力才能將之擊退。侵略國若得逞，對美國的安全與經濟將不利。</p> <p>美國總統警告侵略國：若不撤兵美國將出兵將之擊退。但攻擊國仍繼續侵略。結果美國總統未出兵，攻擊國佔領了鄰國。</p> <p>請問你對美國總統的處理方式，是同意、不同意、還是無所謂同意或不同意？ (若回答同意，追問) 是非常同意，還是普通同意？ (若回答同意，追問) 是非常不同意，還是普通不同意？ (若回答無所謂同意或不同意，追問) 有沒有比較傾向同意或不同意？</p>
--

註 粗體字代表隨機分派之屬性類別。

2.2 聯合實驗 (conjoint experiment)

因應政治現象與議題的多面向特性，政治學近十年來更進一步將「多因子調查」延伸，結合最早在行銷學發展出來的「聯合分析」及計量經濟學中的「表述抉擇實驗」。聯合實驗在屬性類別之隨機組合、以及將組合隨機分派給受訪者評分的设计上，與短文實驗非常類似。但相較於短文式實驗一次只呈現一種情境組合，聯合實驗設計往往一次呈現（表列或短文敘述）兩種或以上之組合選項，請受訪者擇一（稱為 choice-based conjoint；CBC 或 forced-choice conjoint），更能反映受訪者對選項的偏好比較。

政治學者 Hainmueller、Hopkins 及 Yamamoto 等人進一步將主流之「反事實因果推論模型」帶入，改良為「聯合調查實驗」(conjoint survey experiments) (Hainmueller, Hopkins and Yamamoto 2014；Hainmueller and Hopkins 2015；Bansak, Hainmueller, Hopkins and Yamamoto 2021)，頗受政治學界青睞，常應用於探討民眾如何做選擇的研究，例如政策偏好 (policy preferences) 或候選人抉擇 (candidate choice) 等主題。無怪乎 Sniderman (2018) 及 Jenke (2022) 均認為聯合調查實驗，是近年政治學調查實驗最有創意的發展。

聯合調查實驗處理多面向現象的方式如下 (黃紀 2023)：

1. 擇定「因」的若干個屬性及其類別。
2. 決定屬性及其類別的所有可能組合，每一種組合構成一個「選項輪廓」(profile)。
3. 隨機選兩個（或兩個以上）的選項輪廓配成一組 (set)，呈現給受訪者擇一 (forced choice) 或針對各選項評分 (rating)。
4. 重複步驟 3，隨機分派若干回合 (multiple tasks) 給每位受訪者做選擇與評分。

以 Hainmueller and Hopkins (2015) 研究美國民眾對移民 (immigrants) 的偏好為例：

1. 作者擇定移民申請者的 9 個屬性會影響民眾之偏好：性別 (2 類)、教育程度 (7 類)、原國籍 (10 類)、英語能力 (4 類)、專長 (11 類)、工作經歷 (4 類)、就業規劃 (4 類)、曾否到過美國 (5 類)、申請移民的原因 (3 類)。
2. 上述 9 個屬性及其類別共構成 $2 \times 3 \times 4^3 \times 5 \times 7 \times 10 \times 11 = 1,478,400$ 種可能的移民申請者的選項輪廓，但排除不可能的組合（如「申請移民的原因」為逃避迫害者僅限於伊拉克等 3 個原國籍、「專長」之中有 4 種專業至少需有兩年制專科以上的教育程度），實際的組合數減為 $2 \times [(2 \times 7) + (3 \times 3)] \times 4^3 \times 5 \times [(7 \times 7) + (3 \times 4)] = 897,920$ 種。

3. 隨機分派兩個申請移民的輪廓給每位受訪者，先請其就每組兩人之中二擇一准許移民美國，再對此兩人用 1 至 7 的尺度評定其偏好程度（見表 2）。
4. 重複步驟 3，隨機分派 5 回合給每位受訪者做選擇與評分。

以往文獻認為政黨、候選人及社會團體對移民政策眾說紛紜，意見分歧。但 Hainmueller and Hopkins (2015) 將此一隨機分派之聯合實驗，以網調訪問約 1,407 受訪者，分析後得出截然不同的結論：美國民眾其實對移民有高度的共識，幾乎一面倒偏好高較育程度、有一技之長、英語流暢、準備好來美國就業而非依親的移民申請者，他們稱此為「隱性共識」(hidden consensus)。Hainmueller 等人 2014 及 2015 的兩篇論文，不僅刺激了好幾篇以類似之實驗設計，探討其他國家對美國移民政策態度的文章（如 Duch, Laroze, Reinprecht and Robinson 2022），更將聯合實驗迅速擴展到其他政治學的次領域，包含：

1. 比較政治：對候選人的抉擇與偏好（Horiuchi, Smith and Yamamoto 2018；Abramson, Kocak and Magazinnik 2022；Schwarz and Coppock 2022；張佑宗、曾煥凱 2023）、對貪腐政客之究責（Klašnja, Lupu and Tucker 2021）、對民意代表與稅賦的觀感（Kao, Lü and Queralt 2024）、對組聯合內閣對象的課責看法（Liu 2022）、對降低選舉與被選舉權年齡之修憲公投的歧見（Huang 2023）等；
2. 公共政策：對環境政策的支持度（Wicki et al. 2020）、對高齡政策之資源分配的態度（Wu and Kuo 2022）等；
3. 國際關係：民眾對國家「決心」之觀感（Kertzer, Renshon and Yarhi-Milo 2021）、對外交政策之看法（Lim and Tanaka 2022）等。

聯合調查實驗除了可處理更符合實際的多面向複雜議題外，近年也常被應用於敏感議題之民調，如種族、性別平等（Carey, Clayton and Horiuchi 2020）、鄰避效應（Hankinson 2018）等，以降低社會群體應然面壓力產生之偏誤（social desirability bias；SDB）。Horiuchi, Markovich and Yamamoto (2022) 亦證實，聯合實驗設計將多個變數之屬性隨機組合、兩兩配對比較的題型，相較於一般問卷的直接問法（direct questions），可大幅降低敏感題之社會應然偏差。例如黃紀（Huang 2023）將聯合實驗應用至 2022 年 11 月 26 日我國首度修憲複決「18 歲公民權複決案」⁵（投票權年齡自 20 歲降為 18 歲以及被選舉權年齡自 23 歲降為 18 歲）的態度，進而測量選民

⁵修憲公投公民需年滿 20 歲才有投票權，與議題公投之年滿 18 歲不同（見 Huang 2021）。

表 2: 聯合調查實驗問卷中兩個選項之聯合表 (conjoint table) 範例。

請仔細閱讀以下兩個申請移民的案例，你個人覺得哪一位應該優先准許移民美國？

	移民申請人-甲	移民申請人-乙
曾否到過美國	曾拿觀光簽證入境一次	曾拿觀光簽證入境一次
申請移民的原因	與在美國的家人團聚	與在美國的家人團聚
原國籍	墨西哥	伊拉克
英語能力	面試時英語會話流暢	面試時英語會話流暢
專長	保母	教師
工作經歷	一至二年	三至五年
就業規劃	未聘但已與美國雇主面試	到美國後再找工作
教育程度	相當於美國之兩年大專學歷	相當於美國之大專學歷
性別	女	男

[1] 假如你必須從這兩位申請人擇一，請問你會優先選誰？甲____乙____

[2] 如果用 1 到 7 來表示您的看法，1 表示美國絕對不應准許該移民、7 表示絕對應該准許該移民，請問您會給甲多少？____

那乙呢？____

對「選舉權」與「被選舉權」不同年齡組合的真實偏好，以釐清「18 歲公民權複決案」未通過及選民投票抉擇的成因。該複決案因有效同意票未過選舉人總額之半數，故未通過。此一結果，與 2022 年 3 月立法院 109 位出席立委全數贊成通過修憲案、選前中選會舉辦修憲辯論會卻找不到任何機關團體代表反方意見等看似一面倒支持的現象，形成鮮明的對比。為何有效票中高達 47.04% 投不同意票？有三種可能的解釋：一為政黨藍綠陣營對抗形成政治兩極化 (Huang and Kuo 2022)，18 歲公民權修憲案係由執政之民進黨發動，在野之國民黨檯面上雖有所顧忌未明言反對，支持藍營的民眾也可能因社會應然面 (social desirability) 的壓力造成「沉默螺旋」(spiral of silence) (Noelle-Neumann 1974) 而不表態，但在投票時仍投反對票或廢票表達不滿。另一個可能的解釋，則為世代差異：部份年長世代對年輕人的不信任，認為 18 歲的年輕人心智還未成熟，連投票抉擇的思慮都不足，遑論成為候選人。第三個可能的解釋，則為「棄車保帥」：部份選民並不反對投票權年齡自 20 歲降為 18 歲，但反對被選舉權年齡自 23 歲降為 18 歲，惟因複決案僅有一票，只好投反對票。為檢測這三個假設，一方面必須拆解選民對「選舉權」與「被選舉權」的看法，一方面又需避免社會應然面壓力對一般調查之直接問法產生的偏誤，Huang (2023) 採聯合網調實驗設計：

1. 針對 2022 年修憲複決之脈絡，擇定三個屬性：提案支持的政黨 (3 類：民進黨、

國民黨、台灣民眾黨)、投票權年齡(2類:18、20歲)、被選舉權年齡(3類:18、20、23歲)。

2. 上述3個屬性及其類別共構成 $3 \times 2 \times 3 = 18$ 種組合,但因被選舉權年齡只能大於或等於投票權年齡,故排除3種不可能的組合,總共有15種組合的選項輪廓。
3. 就15種組合,隨機分派兩個選項輪廓給每位受訪者,請其圈選支持的選項。
4. 重複步驟3,每位受訪者共填答隨機分派之5回合成對之選項輪廓。

Huang (2023) 針對「台灣政經傳播研究」(黃紀、張卿卿 2023) 的面訪追蹤樣本進行網調,共完成1,134份成功受訪者樣本。

如前所述,聯合網調實驗可針對多個屬性進行因果推論。Hainmueller, Hopkins and Yamamoto (2014) 證明,隨機分派之聯合調查實驗,其因果參數「平均邊際成分效應」(average marginal component effect; AMCE) 之無母數估計,只需以熟悉之多元線性迴歸分析,其係數估計值 AMCE 可解讀為「對選項的中選機率所產生的影響」:亦即將所有其他屬性固定不變於其平均值,因果屬性的某一類相較於其參照類,選項中選機率的增(減)幅度。因為屬性類別為隨機分派,故 AMCE 有明確之因果推論解讀。準此,Huang (2023) 的聯合分析發現:選民對複決案中降低被選舉權年齡有強烈的保留,其他條件不變,被選舉權從23歲降到18歲,受訪者支持修憲案的機率平均驟降約0.395,此一反對的態度遠高於對降低投票權的提議(約0.116)(參見表3之AMCE點估計值及穩健之標準誤、圖3之AMCE點估計及其95%信心區間)。換言之,相當比例投反對票或棄權的選民,其實是對參選公職權從23歲降至18歲的提案感到高度疑慮。

聯合調查實驗雖非政治學者原創,但近十年來在方法上的發展與應用,確實將之發揚光大,使聯合分析與因果推論主流之「潛在後果」架構結合更緊密、推論更嚴謹。自 Hainmueller, Hopkins and Yamamoto (2014) 出版後,引發了政治學界有關聯合實驗設計與分析的豐富文獻(見附錄A),以下舉其中之要點說明:

1. 平均因果效應之解讀: Bansak, Hainmueller, Hopkins and Yamamoto (2023) 指出 AMCE 不僅有明確之因果推論解讀,在選舉研究上亦極為重要,因為 AMCE 估計值摘述了選民多面向偏好的方向與強弱度(directionality and intensity),反映了因果屬性的類別改變後,對候選人或政黨之預期得票率所產生的增減影響。

表 3: 平均邊際成分效應 (AMCE) 之估計值。

	平均邊際成分效應	
	估計值	穩健之標準誤
有無提示憲法 130 條 (參照類: 不提示)		
不提示	-	
有提示	-0.0023	0.0032
提案之政黨 (參照類: 民進黨)		
國民黨	-0.0305*	0.0127
民進黨	-	
台灣民眾黨	0.0148	0.0129
投票權年齡 (參照類: 20)		
18	-0.1155***	0.0134
20	-	
被選舉權年齡 (參照類: 23)		
18	-0.3949***	0.0163
20	-0.1541***	0.0129
23	-	

註 * $p < 0.05$, *** $p < 0.001$ 。

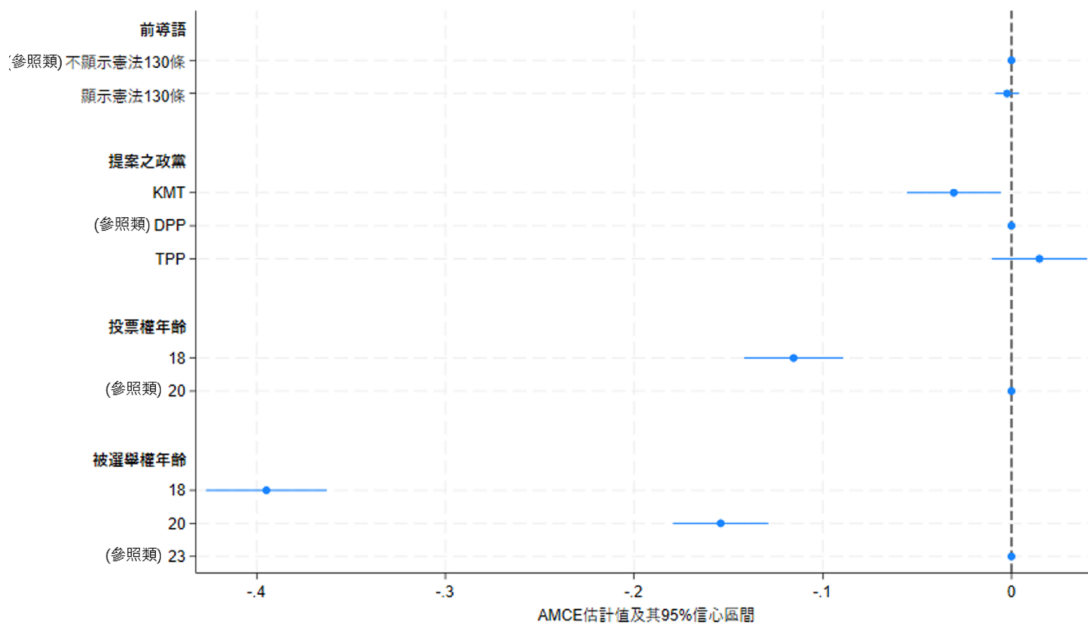


圖 3: 選民「投票權、被選舉權」年齡的看法對贊成修憲複決第一案的因果效應 (各變數參照類之 AMCE 均設為 0)。

不過正因為聯合實驗的特色為同時估計多面向，[Liu and Shiraito \(2023\)](#) 點出多重假設檢定（multiple hypothesis testing）的問題，並提出校正 p 值的方法，以避免錯誤推論。

2. 平均因果交互效應：[Egami and Imai \(2019\)](#) 進一步對屬性間的交互作用，提出「平均因果交互效應」（average marginal interaction effect；AMIE）的估計，使得屬性之因果效應估計更為細緻。例如二維交互效應是指兩個隨機分派的屬性 A 、 B 之類別組合 (a_g, b_h) ，相對於其參照組合 (a_0, b_0) ，在扣除 A 與 B 個別的主效應之後，仍有的因果效應。亦即：

$$\text{AMIE}_{AB} = E \left[Y_i(a_g, b_h) - Y_i(a_0, b_0) - \int \{Y_i(a_g, B_i) - Y_i(a_0, B_i)\} dF(B_i) - \int \{Y_i(A_i, b_h) - Y_i(A_i, b_0)\} dF(A_i) \right]$$

[Ham, Imai and Janson \(2024\)](#) 更導入機器學習（machine learning）的方法，以條件式隨機檢定法（conditional randomization test, CRT）來檢定聯合實驗中「某個屬性在以另一個屬性為條件的情況下，是否仍有因果效應」之假設。

3. 次群體間因果效應之異值性：研究者往往不只對整體的平均因果效應感興趣，還想進一步比較該效應在各個次群體（subgroup）之間是否有差異。直觀上，只需比較各次群的 AMCE 估計值（conditional AMCEs）即可。但 [Leeper, Hobolt and Tilley \(2020\)](#) 及 [Ganter \(2023\)](#) 指出：由於 AMCE 建基於和參照類的比較，而參照類在各次群中的立足點未必相同，使得各次群 AMCE 的相互比較可能有偏誤，因此認為應該估計各次群的邊際平均值（marginal means；MM）及其次群間的差值，以比較各次群對不同選項的偏好（preferences）。在二擇一的聯合實驗中，若以 Y 代表二分類依變數， T 代表各屬性類別的組合， X 代表受訪者之協變數，則協變數的第 r 類，其第 l 個屬性之邊際平均值，就是以 Y 之選擇機率所反映之偏好度：

$$\text{MM}_l = E(Y|T = t_l, X = x_r)$$

由於屬性不論分為幾類，均各估算其邊際平均值，故不會受參照類之影響。

4. 因果調節效應：次群分析固然可以檢驗因果效應是否有異值性，但因次群之「測前特徵」（協變數 covariates）並非隨機分派，因此各次群的邊際平均值並不具備因果效應的解讀。為了符合反事實模型之因果調節效應（causal

moderation)，[Bansak \(2021\)](#) 參照觀察研究中推論因果效應的方法，提出了「平均因果調節效應」(average treatment moderation effect, ATME)。以二分類屬性 T 、二分類調節變數 S 及其他協變數 X 為例，其反事實之潛在後果 $Y(t, s)$ 呈表 4 之 2×2 表：

$$\begin{aligned} \text{ATME} &= E\{Y_i(1, 1) - Y_i(0, 1)\} - \{Y_i(1, 0) - Y_i(0, 0)\} \\ &= E_X [E[Y_i|T_i = 1, S_i = 1, X_i] - E[Y_i|T_i = 0, S_i = 1, X_i] \\ &\quad - E[Y_i|T_i = 1, S_i = 0, X_i] + E[Y_i|T_i = 0, S_i = 0, X_i]] \\ &= E_{X|T=1} [E[Y_i|T_i = 1, S_i = 1, X_i] - E[Y_i|T_i = 1, S_i = 0, X_i]|T_i = 1] \\ &\quad - E_{X|T=0} [E[Y_i|T_i = 0, S_i = 1, X_i] - E[Y_i|T_i = 0, S_i = 0, X_i]|T_i = 0] \end{aligned}$$

由於上式之調節變數 S 並非隨機分派，[Bansak \(2021\)](#) 建議結合觀察研究中推論因果效應的方法，分別在實驗組和控制組之中進行配對 (matching)，例如以傾向分數 (propensity score) 將 $S_i = 1$ 及 $S_i = 0$ 者按照可能會造成干擾之協變數 X_i 進行組內配對，估計各組內 S 對 Y 之因果效應，然後再計算這兩個組內效應之間的差值，即是 ATME 之估計值。ATME 以反事實模型的基本架構，結合實驗與非實驗數據之因果推論方法，為聯合實驗從「描述因果效應之異值性」邁入「推論因果調節效應」奠定了基礎。

3. 聯合實驗之檢討與展望

儘管聯合實驗的發展快速，日臻完備，但仍有些亟待補強之處（參見 [Druckman 2022](#)）。此外，如前所述，政治學在實驗法的發展與應用，相對於其他學科較為晚近，因此其他學科的經驗，都應視為「他山之石」，可供借鏡。

表 4: 二分類屬性與調節變數構成之潛在後果。

		非隨機分派之調節變數 S	
		0	1
隨機分派之屬性 T	0	$Y(0, 0)$	$Y(0, 1)$
	1	$Y(1, 0)$	$Y(1, 1)$

3.1 統計檢定力與樣本數

心理學是社會科學中採實驗法的先進，其 2015 年爆發之實驗「重製危機」(replication crisis, 參見 [Open Science collaboration 2015](#)) 值得政治學深刻警惕。其中涉及之樣本數 (sample size) 與統計檢定力 (statistical power) 問題，與聯合實驗之多面向、多題組結構息息相關，應予正視。因為「二擇一聯合實驗設計」會隨機分派 t 套之配對聯合表給每位受訪者，若將 n 位受訪者的回答疊成長式 (long form) 資料檔，則共有 $n \times t \times 2$ 個觀察值 (observations)。因此常認為聯合實驗的優點之一是執行成本較低，因為只要每份樣本的題組套數夠多，受訪者數不需太大。例如 [Orme and Chrzan \(2017\)](#) 建議計算聯合實驗所需受訪者數的大致法則為 $nta/c \geq 1,000$ (亦即 $n \geq 1000 \cdot c / (t \cdot a)$)， n 為受訪者樣本數、 t 為題組套數、 a 為每個題組的選項數、 c 為屬性分類數 (levels) 中的最大值。但應注意的是，聯合實驗是屬於「同一受試者重複測量設計」(within-subject design)，而同一位受訪者針對多套題組的回答，顯然不會相互獨立，故 $n \times t \times 2$ 個觀察值會予人總觀察值數夠大的錯覺。因此聯合實驗之研究設計，理應比上述之「大致法則」更為細緻。

[Stefanelli and Lukac \(2020\)](#) 以 [Hainmueller, Hopkins and Yamamoto \(2014\)](#) 的「二擇一聯合實驗設計」為架構，參照 [Cohen \(1988\)](#) 經典之統計檢定力以高於 0.8 為原則，針對 AMCE 之估計與檢定，進行大規模之模擬後，提出了以下幾點建議：

1. 聯合調查實驗應在研究設計的階段，就將「事前統計檢定力分析」(a priori power analysis) 視為規劃的重要一環，從因果效應值 (effect size) 的大小、屬性分類數的多寡等因素，推算至少所需之受訪者數與分派給每位受訪者填答之題組套數等，做為實際執行調查實驗時的下限目標。
2. 若依照學理及過往文獻，事前預期之 AMCE 因果效應值越小，則所需之總觀察值數 $n \times t \times 2$ 越大，才能維持 0.8 以上之統計檢定力。但應注意的是，研究者需於設計階段就進行前述「事前統計檢定力分析」，不宜採「事後統計檢定力分析」(post hoc power analysis)，亦即以實驗執行完畢後之 AMCE 樣本估計值回推檢定力 ([Gelman 2019](#); [Liu and Yamamoto 2020](#))。
3. 在屬性的類別完全隨機分派的前提下，屬性的個數對檢定力並無影響，但屬性分成幾類的類別數則有影響，也就是類別數越多，所需之總觀察值數也越大。因此屬性分成哪幾類，研究者需參照學理及研究議題慎重考量。
4. 受訪者數與分派給每位受訪者填答之題組套數之間確實有抵換 (trade-off) 之

關係，在題組套數不會造成受訪者敷衍填答的範圍內 (Bansak, Hainmueller, Hopkins and Yamamoto 2018)，研究者可以提高題組套數 t 來換取較少的受訪者數 n ，或增加受訪者數來換取較少之題組套數。圖 4 係 Stefanelli and Lukac (2020) 設定顯著水準 $\alpha = 0.05$ 、屬性至多分為 5 類、預期最小之 AMCE 因果效應值為 0.03 時，黑色下滑曲線代表維持統計檢定力 0.8 的受訪者數 n (橫軸) 與題組套數 t (縱軸) 之各種組合， n 越大 t 可越小。

5. 綜合上述考量，聯合調查實驗並非如發展初期倡議者所述「低成本、多面向、高效率」那麼樂觀，需要更細緻的精算。一如其他之實證研究法，從學理出發、搭配務實之規劃，做整體之考量，學者責無旁貸。

3.2 從次群異值分析到因果調節效應

前已述及，研究者往往不只關注整體的平均因果效應，還常想進一步比較該效應在各個次群體之間是否有差異。次群間效應之異值性 (treatment effects

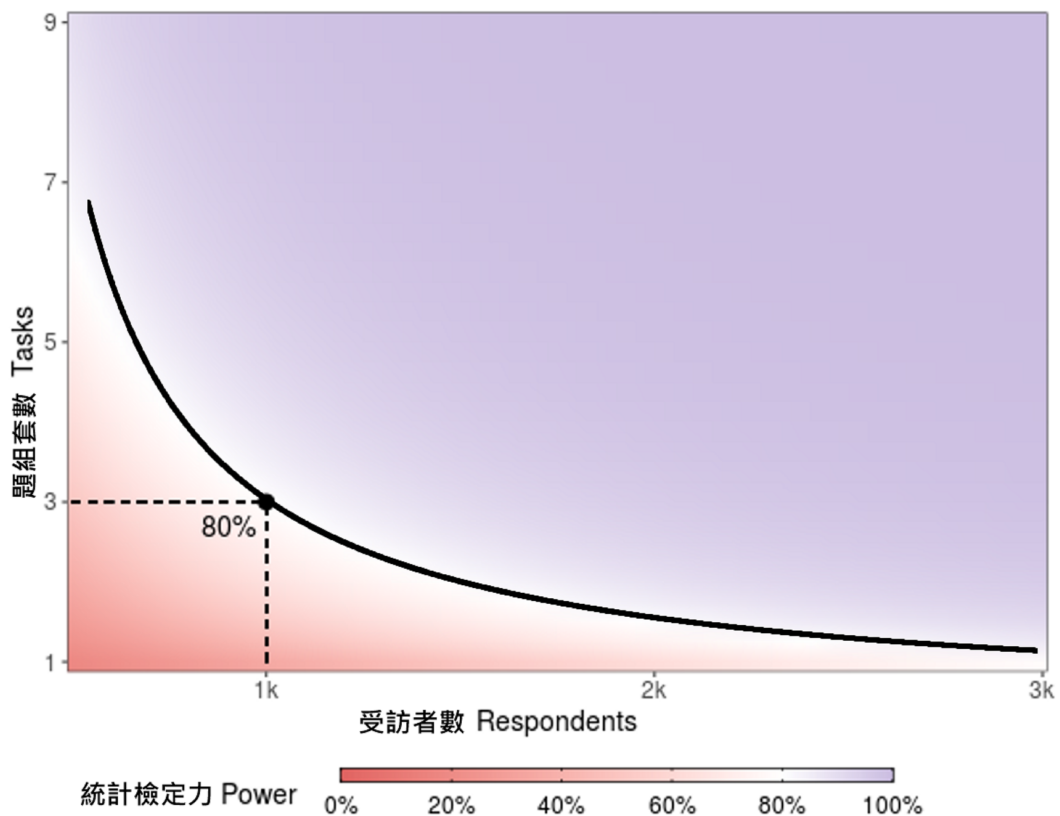


圖 4: 可維持統計檢定力 0.8 的受訪者數 (橫軸) 與題組套數 (縱軸) 間之抵換關係。

heterogeneity)，一直是學界關心的焦點，但對聯合實驗，也構成了以下的挑戰：

1. 總觀察值數的需求增加：政治方法學者 [Gelman \(2018\)](#) 早已指出，解釋變數間之交乘互動，需有更大之樣本數支撐，才有足夠的統計檢定力。此一原則，當然也適用於聯合實驗之分析。因此 [Leeper, Hobolt and Tilley \(2020\)](#) 及 [Ganter \(2023\)](#) 提出之次群體偏好之估計，會增加對總觀察值數的需求。但這方面的具體檢定力分析，文獻尚待著墨。
2. 次群異值分析不等於因果調節效應：聯合實驗之應用者常因屬性類別為隨機分派，便誤以為屬性之因果解讀也可適用於次群異值分析。但因次群之測前協變數 (covariates) 並非隨機分派，因此不論是邊際平均值 (MM) 或平均成分偏好 (average component preference, ACP) 都不具備因果效應的解讀。[Bansak \(2021\)](#) 提出之 ATME 估計，則是將非實驗之觀察研究文獻所發展之配對法等，應用至聯合實驗之協變數，使 ATME 也具有因果效應的解讀。這顯示實證政治學研究者必須具備因果推論的全方位素養，才能跨越隨機實驗與觀察研究，截長補短，交相為用而不劃地自限。

4. 醞釀中之二次革命：外在效度

截至目前，「可信度革命」的重心，在確立因果關係的內在效度 (internal validity)，也就是透過嚴謹的研究設計，強化因與果之間的連結。但在這場革命方興未艾之際，研究者已開始警覺過度執著於內在效度，會導致「明察秋毫之末，而不見輿薪」的盲點，甚至本末倒置，只為應用時興的方法而挑揀偏狹的題材，輕則無關宏旨，忽略了實證研究肩負累積知識與解決重大問題的責任，重則淪為瞎子摸象以偏蓋全，均有違經典之實驗法文獻 ([Campbell and Stanley 1963](#)；[Shadish, Cook and Campbell 2002](#)) 主張因果推論需兼顧外在效度 (external validity) 的初衷，故積極提倡落實因果關係由樣本推廣 (generalizability) 至原母群 (parent population)、乃至將研究發現跨越 (transportability) 適用至不同標的母群 (target population) 的外在效度 (參見 [Coppock 2019](#)；[Findley, Kikuta and Denly 2021](#)；[Hartman 2021](#))。就聯合實驗而言，此一挑戰已逐漸浮現 ([Druckman 2022](#))。

惟外在效度涉及的層面甚廣，顯得眾說紛紜。[Egami and Hartman \(2023\)](#) 綜合相關文獻，提出了一套較為完整的「四面向外在效度架構」：

1. 母群 (populations) 之外在效度 (X-validity)：協變數 X 之分布與母群分布之

符合度；

2. 「因」(treatments) 之外在效度 (T -validity)：原因變數的施測方式；
3. 「果」(outcomes) 之外在效度 (Y -validity)：結果變數的測量與印證；
4. 脈絡 (contexts) 之外在效度 (C -validity)：跨越至不同的時間、區域或制度。

Egami and Hartman (2023) 認為研究者應先依議題及學理鎖定探討的標的面向，並決定要推廣的目標是因果效應的有無與方向 (sign-generalization) 還是更精準的效應幅度之大小 (effect-generalization)，再選擇對應之檢定或提升外在效度的統計方法。若由此一架構觀照聯合實驗的外在效度文獻，可歸納如下：

1. 樣本代表性之調整：儘管 Mutz (2011) 的理想是結合實驗的因果內在效度和調查的代表性樣本之外在效度，可取兩者之長而捨其短。但由於近年網路調查的方便快速與成本較低，許多採網調之聯合實驗，常因無反應或拒訪率偏高，損及樣本代表性，甚至可能是網上招募自願填答 (opt-in) 之方便樣本 (convenience sample)，並非機率樣本。一般直觀的處理方式，往往採非實驗之民調加權方式進行調整，例如反覆加權法 (raking) 等。但畢竟調查實驗除了樣本代表性之外，還有「因」的分組問題，將調查實驗視同一般民調加權，難免有顧此失彼的疑慮。參照 Egami and Hartman (2023) 的架構，重點其實在因果效應是否會隨受訪者之協變數 X 而異，因此就協變數符合原母群分布之的外在效度 (亦即 X -validity) 而言，他們推薦採用觀察研究中因果推論的雙重穩健 (double robust) 估計法，這與前一節 Bansak (2021) 提出之 ATME 估計的思維相互呼應。若進而欲將實驗跨越至其他標的母群，則需同時兼顧母群及脈絡效度 (亦即 X -validity + C -validity) 兩個面向。
2. 選項輪廓之組合分布：聯合實驗在屬性類別之隨機分派上，目前多數應用研究隱含設定選項輪廓之組合為均勻分布 (uniform distribution)。de la Cuesta, Egami, and Imai (2022) 認為均勻分布之假定過度簡化，研究者應該努力蒐集研究標的母群中選項輪廓之實際分布，或是以學理設定之適當分布，做為隨機分派之權重依據，以提升該聯合實驗分析之外在效度。換言之，即是提升「因」之外在效度 (T -validity)。
3. 偏好之測量與外顯行為之驗證：聯合實驗不論採「選項二擇一」或「個別選項尺度評分」，都是測得受測者陳述 (stated) 之選擇與偏好程度。但研究者也常希望此一測量能反映實際之行為。在行銷學中，聯合實驗測得的消費者偏好，

常也有消費行為之實際外顯 (revealed) 資料可資印證，檢證「果」之外在效度 (Y-validity)；政治學的聯合實驗，與真實行為比對的挑戰甚鉅，但並非不可行。Hainmueller, Hangartner, and Yamamoto (2015) 以瑞士之公投為自然實驗，檢證其聯合實驗之預測與公投實際結果之比對，即是一例。

4. 檢驗學理與評估政策的差異：學理的經驗意涵 (empirical implications)，往往僅強調因果關係的有無與正負之方向，因此檢驗學理的實驗，可務實地專注於推廣因果效應的方向。但如果實驗的目的是為了規劃或評估政策，則更需面對推廣因果效應之幅度大小的挑戰。

5. 結論

2000 年之後，實驗法已漸躋身為政治學實證研究的主流之一。本文回顧了本世紀以來因果推論的黃金標準——實驗法——與傳統之調查研究結合而成之「調查實驗」，並聚焦於政治學兩種常見的多因子實驗設計：短文實驗與聯合實驗，說明其研究設計之結構、發展與應用。

短文實驗以一段精簡的文字，嵌入若干個「因」的屬性，以隨機分派的屬性類別組合成假想的標的或情境輪廓，讓受訪者針對該短文表達自己的偏好或評價，據以分析屬性對偏好的影響。短文實驗的優點在於可包含多個屬性，處理多面向的議題，但傳統單一式短文實驗測量較不精準，且難以比較受訪者在屬性之間的權衡。因此近十年來，改進之聯合調查實驗日益受到重視，因其一次呈現兩種組合選項，請受訪者就二者擇一，並隨機分派若干套題組給每位受訪者做選擇，能更精準的測量受訪者的選擇與偏好，且能降低社會應然面壓力造成的偏誤。由於設計上更有彈性，加上近年網路調查在成本及效率上的優勢，聯合調查實驗的應用幾乎已擴及政治學實證研究的各個次領域。在估計與分析方法上，除了與反事實因果推論模型緊密結合的「平均邊際成分效應」外，更進一步延伸至因果效應的異值性，包括屬性間交互作用的「平均邊際交互效應」檢測，以及比較因果效應是否會隨受訪者之協變數特徵而異，分析上更為細緻。

聯合調查實驗的廣泛應用也顯現幾項弱點待加強，值得正視。首先，研究者應在聯合實驗的設計階段，就先以統計檢定力至少維持在 0.8 以上為基礎，考量預期效應之最小值、每位受訪者分派之題組套數、屬性之最大類別數等因素，更精準的預估該實驗至少需完成之有效受訪者樣本數，納入實驗執行之規劃。其次，比較各個次群體之間效應的異值性時，應以不受限於參照類的「邊際平均值」來描述。但若欲推論某

個測前協變數是否決定了屬性因果效應之異值性，則需警惕協變數並非隨機分派，應結合觀察研究之因果推論方法，進行「平均因果調節效應」之估計與分析。

調查實驗發展至今，重心大多集中在確立因果推論的內在效度，已累積相當成果。展望未來，最大的挑戰在推廣因果關係之外在效度。惟調查實驗之外在效度，涉及的層面比非實驗之傳統民調更廣，除了必須顧及協變數分布與母群的代表性外，亦須兼顧因、果、脈絡等的外在效度，亟需學界投入發展整合之架構與分析方法。

附錄 A. 政治學界近年有關「聯合調查實驗」方法論之著作摘述。

作者 (出版年)	聯合調查實驗之設計與分析方法	研究主題	樣本數
Hainmueller, Hopkins and Yamamoto (2014)	聯合調查實驗之設計、平均邊際成分效應 (AMCE) 之估計與因果效應解讀	估計方法	Candidate experiment : 311。(見該文 p. 17) Immigrant experiment : 共執行 2 波定群追蹤調查： 第 1 波：1,714 第 2 波：1,407 詳見 Hainmueller and Hopkins (2015)
Hainmueller, Hangartner, and Yamamoto (2015)	以瑞士公投為自然實驗，比較短文實驗與聯合實驗的外在檢證，發覺成對之聯合實驗 (paired conjoint) 與公投結果最接近，單一短文 (single vignette) 最差	表述偏好與實際投票結果之比對	1,979。(見該文 p. 2396)
Bansak, Hainmueller, Hopkins and Yamamoto (2018)	測試聯合實驗問卷的題組套數 (choice tasks) 會造成受訪者敷衍填答 (satisficing) 的極限，發覺在 30 套以內問題都不大	研究設計	共執行 6 波調查，各波樣本數如下： 第 1 波：605 第 2 波：674 第 3 波：725 第 4 波：1,340 第 5 波：1,577 第 6 波：1,613 (見該文 p. 114)

註 * 本表所列著作，均為「二擇一」式之聯合實驗，故每套題組的選項數均為 2。

** Google scholar 引用數查詢日期：2024 年 04 月 02 日。

† 原作者未提供組合數，筆者自行計算。

‡ 175 頁與 176 頁為同一表格。

題組數 *	屬性數	組合數	搭配之統計軟體	Google Scholar 引用數 **
6	8	$2^2 \times 6^6 = 186,624^\dagger$	R package “cjoint”	1,574
5	9	$2 \times [(2 \times 7) + (3 \times 3)] \times 4^3 \times 5 \times [(7 \times 7) + (3 \times 4)] = 897,920$ (Hainmueller and Hopkins 2015)。		
10	7	$2 \times 3 \times 4^4 \times 8 = 12,288^\dagger$		1,148
30	至多 20	$2^6 \times 3^2 \times 4^1 \times 5 =$ 至少 > 120億。 [†]		209

註 * 本表所列著作，均為「二擇一」式之聯合實驗，故每套題組的選項數均為 2。

** Google scholar 引用數查詢日期：2024 年 04 月 02 日。

† 原作者未提供組合數，筆者自行計算。

‡ 175 頁與 176 頁為同一表格。

作者 (出版年)	聯合調查實驗之設計與分析方法	研究主題	樣本數
Bansak, Hainmueller, Hopkins and Yamamoto (2019)	測試聯合實驗問卷的屬性 (attributes) 數目會造成受訪者敷衍填答 (satisficing) 的極限，發覺可多至 18 個屬性	研究設計	共執行 3 波調查，各波樣本數如下： 第 1 波：1,199 第 2 波：2,476 第 3 波：422 (見該文 p. 62)
Egami and Imai (2019)	針對屬性間的交互作用 (interaction)，提出「平均因果互動效應」(AMIE) 估計因果效應之異值性	估計方法	547，但有 3 個樣本有缺失值，分析時僅投入 544 個樣本。(見該文 p. 530)
Leeper, Hobolt and Tilley (2020)	次群偏好分析及「邊際平均數」(MM) 之估計	估計方法	資料來源取自以下研究：(見該文 p. 23, 24-25) 2,000。資料取自 Ballard-Rosa, Martin and Scheve (2016) 。該資料於原作者稱有 2,000 個正式樣本、500 個前測樣本，但本文作者提供之驗算資料檔內為 2,250 個樣本。 8,500。資料取自 Bechtel, Genovese and Scheve (2017) 之跨國調查，其中法國、德國、英國各有 2,000 樣本，美國則有 2,500 樣本。 311。資料取自 Hainmueller, Hopkins and Yamamoto (2014) 。

註* 本表所列著作，均為「二擇一」式之聯合實驗，故每套題組的選項數均為 2。

** Google scholar 引用數查詢日期：2024 年 04 月 02 日。

† 原作者未提供組合數，筆者自行計算。

‡ 177 頁與 178 頁為同一表格。

題組數 *	屬性數	組合數	搭配之統計軟體	Google Scholar 引用數 **
15	至多 36	$2^5 \times 3^8 \times 4^2 \times 5 =$ 至少 > 18,467 千兆。†		151
3	4	$2^2 \times 3 \times 7 = 84^\dagger$	R package “FindIt”	125
--	---	---	R package “cregg”	525
8	7	$4^4 \times 5^2 \times 6 = 38,400^\dagger$		
4	6	$3^2 \times 4^3 \times 5 = 2,880^\dagger$		
6	8	$2^2 \times 6^6 = 186,624$		

註 * 本表所列著作，均為「二擇一」式之聯合實驗，故每套題組的選項數均為 2。

** Google scholar 引用數查詢日期：2024 年 04 月 02 日。

† 原作者未提供組合數，筆者自行計算。

‡ 177 頁與 178 頁為同一表格。

作者（出版年）	聯合調查實驗之設計與分析方法	研究主題	樣本數
Leeper, Hobolt and Tilley (2020)	次群偏好分析及「邊際平均數」(MM)之估計	估計方法	1,470。關於該資料集說明請見 Hainmueller and Hopkins (2015) 。但本文作者提供之驗算資料檔內為 1,396 個樣本。
			5,088。資料取自 Teele, Kalla and Rosenbluth (2018) 。該研究稱共執行三項調查，第 1 項調查對象為美國地方與中央層級代議員共 1,866 樣本；第 2 項調查為合格選民共 2,144 樣本；第 3 項調查為美國地方層級代議員共 1,078 樣本；但本文作者提供之驗算資料檔內為第 1 項調查共 1,867 個樣本；第 2 項調查共 2,105 個樣本；第 3 項調查共 312 個樣本。
Stefanelli and Lukac (2020)	聯合實驗設計之統計檢定力分析 (power analysis)	統計檢定	模擬資料：模擬 500 至 3,000 個觀測值，模擬 1,000 次。(見該文 p. 13)
Bansak (2021)	針對屬性與協變數間的因果調節作用 (causal moderation)，提出「平均因果調節效應」(ATME) 的估計	估計方法	資料來源取自 Bansak, Hainmueller and Hangartner (2016) 於奧地利等歐洲 15 個國家執行調查共 18,030 個樣本，各國家別與各國完成之樣本數列表請參閱該文補充文件 p. 7。
Zhirkov (2022)	若每位受訪者也對多個選項做評分 (numerical rating)，則可估計「個體邊際成分效應」(individual marginal component effects；IMCE)	估計方法	929。該調查共完成 1,003 個樣本，但其中 74 個樣本因各題測量值完全相同而刪除，分析時僅投入 929 個樣本。(見該文 p. 241)

註 * 本表所列著作，均為「二擇一」式之聯合實驗，故每套題組的選項數均為 2。

** Google scholar 引用數查詢日期：2024 年 04 月 02 日。

† 原作者未提供組合數，筆者自行計算。

‡ 179 頁與 180 頁為同一表格。

題組數 *	屬性數	組合數	搭配之統計軟體	Google Scholar 引用數 **
5	9	$2 \times [(2 \times 7) + (3 \times 3)] \times 4^3 \times 5 \times [(7 \times 7) + (3 \times 4)] = 897,920$ (Hainmueller and Hopkins 2015)。		
3	6	$2 \times 3^3 \times 4^2 = 864^\dagger$		
1 至 9	5	每個特徵 2 至 20 種等級，至多 320 萬。 [†]	網站： https://mblukac.shinyapps.io/conjoints-power-shiny/	31
5	9	$2 \times 3^4 \times 4 \times 5 \times 6 \times 7 = 136,080^\dagger$		43
15	6	$2^2 \times 4^3 \times 6 = 1,536^\dagger$		22

註 * 本表所列著作，均為「二擇一」式之聯合實驗，故每套題組的選項數均為 2。

** Google scholar 引用數查詢日期：2024 年 04 月 02 日。

† 原作者未提供組合數，筆者自行計算。

‡ 179 頁與 180 頁為同一表格。

作者（出版年）	聯合調查實驗之設計與分析方法	研究主題	樣本數
de la Cuesta, Egami, and Imai (2022)	在實驗設計中納入選項輪廓之母群分布，以提升 AMCE 因果推論之外在效度	研究設計	資料來源取自以下研究：（見該文 p. 23, 24-25）
			1,583。資料取自 Ono and Burden (2019) 執行之聯合調查。該調查共先詢問 3,152 位受訪者基本資訊，再邀請其中 1,744 位受訪者參與調查，最終完成 1,583 份聯合調查問卷。
Ganter (2023)	測量偏好（而非選擇機率）之「平均成分偏好」（average component preference；ACP）之估計	估計方法	1,509。資料取自 Peterson (2017)。該調查稱樣本數為 1,507 (Peterson 2017)，但文內各圖所示樣本數皆為 1,059 (見 Peterson 2017, fig. 2-4, pp. 1196-1198)。
			資料來源取自以下研究：（見該文 pp. 109-110，及補充文件 p. 13）
			模擬資料：每次模擬 5,000 個觀測值，共模擬生成 5,000 個資料集。（見該文補充文件 p. 13）
			至多 1,407。資料取自 Hainmueller and Hopkins (2015)，但僅投入部分樣本進行分析。
			至多 4,778。資料取自 Mummolo and Nall (2017)，補充文件表 2[p. 2]，但僅投入部分樣本進行分析。

註 * 本表所列著作，均為「二擇一」式之聯合實驗，故每套題組的選項數均為 2。

** Google scholar 引用數查詢日期：2024 年 04 月 02 日。

† 原作者未提供組合數，筆者自行計算。

‡ 181 頁與 182 頁為同一表格。

題組數 *	屬性數	組合數	搭配之統計軟體	Google Scholar 引用數 **
--	--	--	R package “factorEx”	110
10	13	$2^4 \times 3^2 \times 4^3 \times 5 \times 6^3 = 9,953,280$		
3	10	$2^3 \times 3^3 \times 4 \times 5 \times 6 \times 10 = 259,200^\dagger$		
--	--	--		24
1	4	$\infty \times 2 \times 4^2$ (第一個屬性為連續變數)		
5	9	$2 \times [(2 \times 7) + (3 \times 3)] \times 4^3 \times 5 \times [(7 \times 7) + (3 \times 4)] = 897,920$ (Hainmueller and Hopkins 2015)。		
5	7	$2^2 \times 3^2 \times 4^2 \times 6 = 3,456^\dagger$		

註 * 本表所列著作，均為「二擇一」式之聯合實驗，故每套題組的選項數均為 2。

** Google scholar 引用數查詢日期：2024 年 04 月 02 日。

† 原作者未提供組合數，筆者自行計算。

‡ 181 頁與 182 頁為同一表格。

作者（出版年）	聯合調查實驗之設計與分析方法	研究主題	樣本數
Liu and Shiraito (2023)	針對多重假設檢定（multiple hypothesis testing）的問題，提出校正 p 值的方法	統計檢定	資料來源取自以下研究：（見該文 p. 382, pp. 389-390，及補充文件 p. 2, 5）
			311。資料取自 Hainmueller, Hopkins and Yamamoto (2014) 並於資料集內模擬分析變數。
			2,320。資料取自 Spilker, Bernauer and Umaña (2016) (p.711 及 Appendix 1)。
Bansak, Hainmueller, Hopkins and Yamamoto (2023)	聯合調查實驗之設計、平均邊際成分效應（AMCE）之估計與因果效應解讀	估計方法	模擬資料：每次模擬 1,001 個觀測值，共模擬生成 1,000 個資料集。（見該文補充文件 p. 13）
Ham, Imai and Janson (2024)	以條件式隨機檢定法（conditional randomization test；CRT）來檢定聯合實驗中「某個屬性在以另一個屬性為條件的情況下，是否仍有因果效應」之假設	研究設計	資料來源取自以下研究：（見該文 p. 13）
			1,470。資料取自 Hainmueller and Hopkins (2015)。
			1,583。資料取自 Ono and Burden (2019)。

註* 本表所列著作，均為「二擇一」式之聯合實驗，故每套題組的選項數均為 2。

** Google scholar 引用數查詢日期：2024 年 04 月 02 日。

† 原作者未提供組合數，筆者自行計算。

†† 183 頁與 184 頁為同一表格。

題組數 *	屬性數	組合數	搭配之統計軟體	Google Scholar 引用數 **
---	---	---		8
6	8	$2^2 \times 6^6 = 186,624$		
5	9	$2^2 \times 3^7 = 8,748^\dagger$		
10	8	$2^4 \times 3^2 \times 4^2 = 2,304$	R package “cjoint”	87
---	---	---	R package “CRTConjoint”	5
5	9	$2 \times [(2 \times 7) + (3 \times 3)] \times 4^3 \times 5 \times [(7 \times 7) + (3 \times 4)] = 897,920$ (Hainmueller and Hopkins 2015)。		
10	13	$2^4 \times 3^2 \times 4^3 \times 5 \times 6^3 = 9,953,280$		

註 * 本表所列著作，均為「二擇一」式之聯合實驗，故每套題組的選項數均為 2。

** Google scholar 引用數查詢日期：2024 年 04 月 02 日。

† 原作者未提供組合數，筆者自行計算。

‡ 183 頁與 184 頁為同一表格。

參考文獻

- [1] Abramson, S. F., Koçak, K., and Magazinnik, A. (2022). What Do We Learn about Voter Preferences from Conjoint Experiments? *American Journal of Political Science*, 66(4), pages 1008-1020.
- [2] Angrist, J. D., and Pischke, J. S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of economic perspectives*, 24(2), pages 3-30.
- [3] Ashworth, S., Berry, C. R., and de Mesquita, E. B. (2021). *Theory and Credibility: Integrating Theoretical and Empirical Social Science*. Princeton University Press, United States.
- [4] Atzmüller, C., and Steiner, P. M. (2010). Experimental Vignette Studies in Survey Research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 6(3), pages 128-138.
- [5] Auspurg, K., and Hinz, T. (2015). *Factorial Survey Experiments*. Sage, United States.
- [6] Ballard-Rosa, C., Martin, L., and Scheve, K. (2016). The Structure of American Income Tax Policy Preferences. *The Journal of Politics*, 79(1), pages 1-16.
- [7] Bansak, K. (2020). Comparative Causal Mediation and Relaxing the Assumption of No Mediator–Outcome Confounding: An Application to International Law and Audience Costs. *Political Analysis*, 28(2), pages 222-243.
- [8] Bansak, K. (2021). Estimating Causal Moderation Effects with Randomized Treatments and Non-Randomized Moderators. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 184(1), pages 65-86.
- [9] Bansak, K., Hainmueller, J., and Hangartner, D. (2016). How Economic, Humanitarian, and Religious Concerns Shape European Attitudes toward Asylum Seekers. *Science*, 354(6309), pages 217-222.

-
- [10] Bansak, K., Hainmueller, J., Hopkins, D. J., and Yamamoto, T. (2018). The Number of Choice Tasks and Survey Satisficing in Conjoint Experiments. *Political Analysis*, 26(1), pages 112-119.
- [11] Bansak, K., Hainmueller, J., Hopkins, D. J., and Yamamoto, T. (2019). Beyond the Breaking Point? Survey Satisficing in Conjoint Experiments. *Political Science Research and Methods*, 9(1), pages 53-71.
- [12] Bansak, K., Hainmueller, J., Hopkins, D. J., and Yamamoto, T. (2021). Conjoint Survey Experiments. In: Druckman, J. N., and Green, D. P. (2021). *Advances in Experimental Political Science*. Cambridge University Press, United Kingdom. pages 19-41.
- [13] Bansak, K., Hainmueller, J., Hopkins, D. J., and Yamamoto, T. (2023). Using Conjoint Experiments to Analyze Election Outcomes: The Essential Role of the Average Marginal Component Effect. *Political Analysis*, 31(4), pages 1-19.
- [14] Bechtel, M. M., and Scheve, K. F. (2013). Mass Support for Global Climate Agreements Depends on Institutional Design. *Proceedings of the National Academy of Sciences*, 110(34), pages 13763-13768.
- [15] Bechtel, M. M., Genovese, F., and Scheve, K. F. (2017). Interests, Norms and Support for the Provision of Global Public Goods: The Case of Climate Co-operation. *British Journal of Political Science*, 49(4), pages 1333-1355.
- [16] Blackman, A. D. (2018). Religion and Foreign Aid. *Politics and Religion*, 11(3), pages 522-552.
- [17] Blair, G., Coppock, A., and Moor, M. (2020). When to Worry about Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments. *American Political Science Review*, 114(4), pages 1297-1315.
- [18] Bol, D. (2018). Putting Politics in the Lab: A Review of Lab Experiments in Political Science. *Government and Opposition*, 54(1), pages 167-190.
- [19] Box-Steffensmeier, J. M., Brady, H. E., and Collier, D. (2008). Political Science Methodology. In: Box-Steffensmeier, J. M., Brady, H. E., and Collier, D. (2008).

- The Oxford Handbook of Political Methodology*. Oxford University Press, United Kingdom. pages 3-32.
- [20] Campbell, D. T., and Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Rand McNally College Publishing, United States.
- [21] Carey, J. M., Clayton, K., and Horiuchi, Y. (2020). *Campus Diversity: The Hidden Consensus*. Cambridge University Press, United Kingdom.
- [22] Carreras, M., and Vera, S. (2018). Do Corrupt Politicians Mobilize or Demobilize Voters? A Vignette Experiment in Colombia. *Latin American Politics and Society*, 60(3), pages 77-95.
- [23] Clayton, K., Horiuchi, Y., Kaufman, A. R., King, G., Komisarchik, M. (2023). Correcting Measurement Error Bias in Conjoint Survey Experiments. Working paper.
- [24] Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge, United Kingdom.
- [25] Coppock, A. (2019). Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach. *Political Science Research and Methods*, 7(3), pages 613-628.
- [26] de la Cuesta, B., Egami, N., and Imai, K. (2022). Improving the External Validity of Conjoint Analysis: The Essential Role of Profile Distribution. *Political Analysis*, 30(1), pages 19-45.
- [27] Devaux, M., and Egami, N. (2022). Quantifying Robustness to External Validity Bias. URL: https://naokiegami.com/paper/external_robust.pdf
- [28] Druckman, J. N. (2022). *Experimental Thinking: A Primer on Social Science Experiments*. Cambridge University Press, United Kingdom.
- [29] Druckman, J. N. and Green, D. P. (2021). A New Era of Experimental Political Science. In: Druckman, J. N., and Green, D. P. (2021). *Advances in Experimental Political Science*. Cambridge University Press, United Kingdom. pages 1-16.

-
- [30] Druckman, J. N., and Green, D. P. (2021). *Advances in Experimental Political Science*. Cambridge University Press, United Kingdom.
- [31] Druckman, J. N., Green, D. P., Kuklinski, J. H., and Lupia, A. (2006). The Growth and Development of Experimental Research in Political Science. *American Political Science Review*, 100(4), pages 627-635.
- [32] Druckman, J. N., Green, D. P., Kuklinski, J. H., and Lupia, A. (2011). *Cambridge Handbook of Experimental Political Science*. Cambridge University Press, United Kingdom.
- [33] Duch, R. M., Laroze, D., Reinprecht, C., and Robinson, T. S. (2022). Nativist Policy: The Comparative Effects of Trumpian Politics on Migration Decisions. *Political Science Research and Methods*, 10(1), pages 171-187.
- [34] Egami, N., and Hartman, E. (2023). Elements of External Validity: Framework, Design, and Analysis. *American Political Science Review*, 117(3), pages 1070-1088.
- [35] Egami, N., and Imai, K. (2019). Causal Interaction in Factorial Experiments: Application to Conjoint Analysis. *Journal of the American Statistical Association*, 114(526), pages 529-540.
- [36] Findley, M. G., Kikuta, K., and Denly, M. (2021). External Validity. *Annual Review of Political Science*, 24, pages 365-393.
- [37] Frank, K. A., Lin, Q., Xu, R., Maroulis, S., and Mueller, A. (2023). Quantifying the Robustness of Causal Inferences: Sensitivity Analysis for Pragmatic Social Science. *Social Science Research*, 110.
- [38] Ganter, F. (2023). Identification of Preferences in Forced-Choice Conjoint Experiments: Reassessing the Quantity of Interest. *Political Analysis*, 31(1), pages 98-112.
- [39] Gelman, A. (2018). You Need 16 Times the Sample Size to Estimate an Interaction than to Estimate a Main Effect. URL: <https://statmodeling.stat.columbia.edu/2018/03/15/need16/>

-
- [40] Gelman, A. (2019). Don't Calculate Post-hoc Power Using Observed Estimate of Effect Size. *Annals of Surgery*, 269(1), pages e9-e10.
- [41] Gerber, A. S., and Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton & Company, United States.
- [42] Gerber, A. S., and Green, D. P. (2000). The Effects of Personal Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment. *American Political Science Review*, 94(3), pages 653-663.
- [43] Granato, J., Lo, M., and Wong, M. C. S. (2021). *Empirical Implications of Theoretical Models in Political Science*. Cambridge University Press, United Kingdom.
- [44] Green, D. P. (2022). *Social Science Experiments: A Hands-on Introduction*. Cambridge University Press, United Kingdom.
- [45] Green, P. E., and Rao, V. R. (1971). Conjoint Measurement for Quantifying Judgmental Data. *Journal of Marketing Research*, 8(3), pages 355-363.
- [46] Hainmueller, J., and Hopkins, D. J. (2015). The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes toward Immigrants. *American Journal of Political Science*, 59(3), pages 529-548.
- [47] Hainmueller, J., Hangartner, D., and Yamamoto, T. (2015). Validating Vignette and Conjoint Survey Experiments Against Real-world Behavior. *Proceedings of the National Academy of Sciences*, 112(8), pages 2395-2400.
- [48] Hainmueller, J., Hopkins, D. J., and Yamamoto, T. (2014). Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments. *Political Analysis*, 22(1), pages 1-30.
- [49] Ham, D. W., Imai, K., and Janson, L. (2024). Using Machine Learning to Test Causal Hypotheses in Conjoint Analysis. *Political Analysis*, pages 1-16.
- [50] Hankinson, M. (2018). When Do Renters Behave Like Homeowners? High Rent, Price Anxiety, and NIMBYism. *American Political Science Review*, 112(3), pages 473-493.

-
- [51] Hartman, E. (2021). Generalizing Experimental Results. In: Druckman, J. N., and Green, D. P. (2021). *Advances in Experimental Political Science*. Cambridge University Press, United Kingdom. pages 385-410.
- [52] Hensher, D. A, Rose, J. M., and Greene, W. H. (2015). *Applied Choice Analysis*, Second Edition. Cambridge University Press, United Kingdom.
- [53] Horiuchi, Y., Markovich Z., and Yamamoto, T. (2022). Does Conjoint Analysis Mitigate Social Desirability Bias? *Political Analysis*, 30(4), pages 535-549.
- [54] Horiuchi, Y., Smith, D. M., and Yamamoto, T. (2018). Measuring Voters' Multidimensional Policy Preferences with Conjoint Analysis: Application to Japan's 2014 Election. *Political Analysis*, 26(2), pages 190-209.
- [55] Huang, C. (2015). Endogenous Regressors in Nonlinear Probability Models: A Generalized Structural Equations Modeling Approach. *Journal of Electoral Studies*, 22(1), pages 1-33.
- [56] Huang, C. (2021). Youth Turnout in Referendums and Elections: Evidence from Regression Discontinuity Designs. *Taiwanese Political Science Review*, 25(2), pages 169-218.
- [57] Huang, C. (2023). Defying the Global Trend of Lowering Eligibility Age? A Conjoint Experiment on Taiwan's 2022 Referendum. *2023 Annual Meeting of American Political Science Association (APSA)*. Los Angeles.
- [58] Huang, C., and Kuo, T. C. (2022). Actual and Perceived Polarization on Independence-Unification Views in Taiwan. *Asian Journal of Communication*, 32(2), pages 75-92.
- [59] Huang, C., Kuo, M. F., and Stockton, H. (2016). The Consequences of MMM on Party Systems. In: Batto, N. F., Huang, C., Tan, A. C., and Cox, G. W. (2016). *Mixed-Member Electoral Systems in Constitutional Context: Taiwan, Japan, and Beyond*. University of Michigan Press, United States. pages 25-51.

- [60] Huang, C., Wang, H. C., and Lin, C. C. (2013). Knowledge of the Electoral System and Voting: Taiwan's 2008 and 2012 Legislative Elections. *Issues & Studies*, 49(4), pages 1-45.
- [61] Hyde, S. D. (2015). Experiments in International Relations: Lab, Survey, and Field. *Annual Review of Political Science*, 18, pages 403-424.
- [62] Imai, K., Keele, L., Tingley, D., and Yamamoto, T. (2011). Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *American Political Science Review*, 105(4), pages 765-789.
- [63] Imai, K., King, G., and Stuart, E. A. (2008). Misunderstandings between Experimentalists and Observationalists about Causal Inference. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 171(2), pages 481-502.
- [64] Imbens, G. W., and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, United Kingdom.
- [65] Jenke, L. (2022). Introduction to the Special Issue: Innovations and Current Challenges in Experimental Methods. *Political Analysis*, 30(S1), pages S3-S7.
- [66] Kao, J. C., Lü, X., and Queralt, D. (2024). Do Gains in Political Representation Sweeten Tax Reform in China? It Depends on Who You Ask. *Political Science Research and Methods*, 12(1), pages 146-165.
- [67] Kaufman, E. (2019). Can Narratives of White Identity Reduce Opposition to Immigration and Support for Hard Brexit? A Survey Experiment. *Political Studies*, 67(1), pages 31-46.
- [68] Kertzer, J. D., Renshon, J., and Yarhi-Milo, K. (2021). How Do Observers Assess Resolve? *British Journal of Political Science*, 51(1), pages 308-330.
- [69] Kinder, D. R., and Palfrey, T. R. (1993). On Behalf of an Experimental Political Science. In: Kinder, D. R., and Palfrey, T. R. (1993). *Experimental Foundations of Political Science*. University of Michigan Press, United States. pages 1-39.

- [70] Klašnja, M., Lupu, N., and Tucker, J. A. (2021). When Do Voters Sanction Corrupt Politicians? *Journal of Experimental Political Science*, 8(2), pages 161-171.
- [71] Leeper, T. J., Hobolt, S. B., and Tilley, J. (2020). Measuring Subgroup Preferences in Conjoint Experiments. *Political Analysis*, 28(2), pages 207-221.
- [72] Lim, S., and Tanaka, S. (2022). Why Costly Rivalry Disputes Persist: A Paired Conjoint Experiment in Japan and South Korea. *International Studies Quarterly*, 66(4).
- [73] Liu, G., and Shiraito, Y. (2023). Multiple Hypothesis Testing in Conjoint Analysis. *Political Analysis*, 31(3), pages 380-395.
- [74] Liu, S., and Yamamoto, T. (2020). How Much Should You Trust Your Power Calculation Results? Power Analysis as an Estimation Problem. URL: <https://web.mit.edu/teppeii/www/research/pilotpower.pdf>
- [75] Liu, T. P. (2022). Consider Your Companions Carefully: How Voters Perceive (Coalition) Governments' poor valence images. *Party Politics*, 28(6), pages 1150-1163.
- [76] Louviere, J. J., Hensher, D. A., and Swait, J. D. (2000). *Stated Choice Methods: Analysis and Applications*. Cambridge University Press, United Kingdom.
- [77] Mize, T. D., and Manago, B. (2022). The Past, Present, and Future of Experimental Methods in the Social Sciences. *Social Science Research*, 108.
- [78] Morton, R. B., and Williams, K. C. (2010). *Experimental Political Science and the Study of Causality: From Nature to the Lab*. Cambridge University Press, United Kingdom.
- [79] Mummolo, J., and Nall, C. (2017). Why Partisans Do Not Sort: The Constraints on Political Segregation. *The Journal of Politics*, 79(1), pages 45-59.
- [80] Mutz, D. C. (2011). *Population-Based Survey Experiments*. Princeton University Press, United States.

-
- [81] Noelle-Neumann, E. (1974). The Spiral of Silence: A Theory of Public Opinion. *Journal of Communication*, 24(2), pages 43-51.
- [82] Ono, Y., and Burden, B. C. (2019). The Contingent Effects of Candidate Sex on Voter Choice. *Political Behavior*, 41(3), pages 583-607.
- [83] Open Science Collaboration. (2015). Estimating the Reproducibility of Psychological Science. *Science*, 349(6251).
- [84] Orme, B. K., and Chrzan, K. (2017). *Becoming an Expert in Conjoint Analysis: Choice Modeling for Pros.* Sawtooth Software, Inc., United States.
- [85] Peterson, E. (2017). The Role of the Information Environment in Partisan Voting. *The Journal of Politics*, 79(4), pages 1191-1204.
- [86] Rossi, P. H. (1979). Vignette Analysis: Uncovering the Normative Structure of Complex Judgments. In: Merton, R. K., Coleman, J. S., and Rossi, P. H. (1979). *Qualitative and Quantitative Social Research: Papers in Honor of Paul F. Lazarsfeld.* The Free Press, United States. pages 176-186.
- [87] Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5), pages 688-701.
- [88] Schwarz, S., and Coppock, A. (2022). What Have We Learned about Gender from Candidate Choice Experiments? A Meta-Analysis of Sixty-Seven Factorial Survey Experiments. *The Journal of Politics*, 84(2), pages 655-668.
- [89] Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Houghton Mifflin, United States.
- [90] Sniderman, P. M. (2018). Some Advances in the Design of Survey Experiments. *Annual Review of Political Science*, 21, pages 259-275.
- [91] Spilker, G., Bernauer, T., and Umaña, V. (2016). Selecting Partner Countries for Preferential Trade Agreements: Experimental Evidence from Costa Rica, Nicaragua, and Vietnam. *International Studies Quarterly*, 60(4), pages 706-718.

- [92] Stefanelli, A., and Lukac, M. (2020). Subjects, Trials, and Levels: Statistical Power in Conjoint Experiments. URL: <https://doi.org/10.31235/osf.io/spkcy>
- [93] Teele, D. L., Kalla, J., and Rosenbluth, F. (2018). The Ties That Double Bind: Social Roles and Women's Underrepresentation in Politics. *American Political Science Review*, 112(3), pages 525-541.
- [94] Thomas, K. (2024). The Advent of Survey Experiments in Politics and International Relations. *Government and Opposition*, 59(1), pages 297-320.
- [95] Tomz, M. (2007). Domestic Audience Costs in International Relations: An Experimental Approach. *International Organization*, 61(4), pages 821-840.
- [96] Wallander, L. (2009). 25 Years of Factorial Surveys in Sociology: A review. *Social Science Research*, 38(3), pages 505-520.
- [97] Wicki, M., Huber, R. A., and Bernauer, T. (2020). Can Policy-Packaging Increase Public Support for Costly Policies? Insights from a Choice Experiment on Policies against Vehicle Emissions. *Journal of Public Policy*, 40(4), pages 599-625.
- [98] Wu, J. E., and Kuo, M. F. (2022). How Conjoint Analysis Contributes to Mitigating Intergroup Conflict? A Case Study of Ageing Policy in Taipei City (Taiwan). *Journal of Asian Public Policy*, pages 1-22.
- [99] Zhirkov, K. (2022). Estimating and Using Individual Marginal Component Effects from Conjoint Experiments. *Political Analysis*, 30(2), pages 236-249.
- [100] 郭銘傑 (2022)。不完全訊息與戰爭爆發：一個隨機實驗。《台灣政治學刊》，26卷 1 期，頁 57-96。
- [101] 黃紀 (2008)。因果推論與觀察研究：「反事實模型」之思考。《社會科學論叢》，2 卷 1 期，頁 1-21。
- [102] 黃紀 (2010)。因果推論與效應評估：區段識別法及其於「選制效應」之應用。《選舉研究》，17 卷 2 期，頁 103-134。
- [103] 黃紀 (2013)。政治學計量方法的回顧與前瞻。吳玉山、林繼文、冷則剛 (2013)。《政治學的回顧與前瞻》，臺北市：五南圖書出版股份有限公司。

- [104] 黃紀 (2017)。TEDS 網路調查實驗平台第一次測試報告。《2016 年至 2020 年「臺灣選舉與民主化調查」四年期研究規劃 (1/4)》(MOST 105-2420-H-004 -015-SS4)。
- [105] 黃紀 (2023)。調查研究設計。陳陸輝 (2023)。《民意調查》，臺北市：五南圖書出版股份有限公司。
- [106] 黃紀、張卿卿 (2023)。台灣政經傳播研究：政治極化之定群追蹤調查 (TIGCR-PPS 2018-2022)。國立政治大學台灣政經傳播研究中心。
- [107] 黃紀、林正義、蕭高彥、陳敦源、黃長玲、廖達琪 (2007)。政治學門「熱門及前瞻研究議題」之規劃。《人文與社會科學簡訊》，9 卷 1 期，頁 72-80。
- [108] 張佑宗、曾煥凱 (2023)。網路輿情與調查研究方法的新挑戰。吳重禮、吳文欽、張廖年仲 (2023)。《政治學的現況與展望》，臺北市：五南圖書出版股份有限公司。
- [109] 楊光、鄭栞尹 (2019)。列項實驗與網路調查之結合。《選舉研究》，26 卷 2 期，頁 23-52。

[Received November 2023; accepted June 2024.]

Development and Application of Causal Inference in Political Science: A Case Study of Survey Experiments

Chi Huang[†]

Department of Political Science, National Chengchi University

ABSTRACT

The “counterfactual model of causal inference” has sparked a “credibility revolution” in social science research, and political science is no exception; its development has been rapid. The purpose of this article is to review the “gold standard” of causal inference—experimental methods—and their integration with traditional survey methods to form “survey experiments”. Although political science had a slower start in the development and application of experimental methods, after the year 2000, experimental methods began to emerge as one of the mainstream approaches in empirical political research. Post-2010, it experienced explosive growth, with survey experiments standing out. This article focuses on the past decade, exploring two types of experimental designs in political science survey experiments that deal with multiple factors: vignette experiments and conjoint experiments. It elucidates their research design structures and applications. By examining the design and analysis of conjoint survey experiments, which have seen the fastest development and application in the past decade, it is argued that the first phase of the credibility revolution has been overwhelmingly successful in establishing the internal validity of causal inferences. Looking ahead to future development and application, it is imperative to build on this success and strive to break through in the external validity of causal relationships.

Key words and phrases: Causal Inference, Empirical Political Research, Survey Experiments, Vignette Experiments, Conjoint Experiments.

JEL classification: C90.

[†]Corresponding to: Chi Huang
E-mail: chihuang@nccu.edu.tw