

科學研究的 replication 的對談

「開放咖啡角」網站 Podcast 檔案

[https://open-](https://open-cafe.netlify.app/project/ep010/?fbclid=IwAR2s9GmsGS2MR4YipTFgUGjQqPeCNZFDjHBVHSR49-LOh54EqleOctCPrcs)

[cafe.netlify.app/project/ep010/?fbclid=IwAR2s9GmsGS2MR4YipTFgUGjQqPeCNZFDjHBVHSR49-LOh54EqleOctCPrcs](https://open-cafe.netlify.app/project/ep010/?fbclid=IwAR2s9GmsGS2MR4YipTFgUGjQqPeCNZFDjHBVHSR49-LOh54EqleOctCPrcs)

慈濟大學心理系陳紹慶與中正大學哲學系陳瑞麟

大家好，歡迎收聽開放咖啡角，我是本集主持人陳紹慶。這個 podcast 好不容易來到第 10 集，所以我想要來談談開始這個節目最重要的課題：What is replication?

這個課題也是最近發表在 PLOS Biology 的評論文章標題，作者是開放科學中心 CEO Brian Nosek 教授。這篇評論主張科學家應該將 replication 當成診斷理論品質的工具，他的看法和今年初我在與陳瑞麟老師的網路筆談有許多值得交互參考的地方。因為陳瑞麟老師有聽力障礙，不方便進行即席交談，所以最後我決定整合兩個文本，向聽眾做個介紹。

談到科學研究的 replication，一開始我們想到的應該是後續的研究採用相同的方法與設備材料，能完整重現相同結果，就是成功的 replication。然而 Brian Nosek 寫的評論指出這樣的看法無法促進科學進步，最近有個新聞可以說明 Nosek 教授主張的理由：澳洲政府為了抗新冠肺炎，決定讓前線醫護施打卡介苗，因為有個相關性研究指出全民施打卡介苗的國家，新冠肺炎患者的死亡率顯著低於沒有全民施打卡介苗的歐美國家，而相關性研究的結論通常只能當成有測試價值的假設。疫情嚴峻的時刻，如果還要仔細證實接種卡介苗的人確實有較強的免疫力抵抗肺炎病毒，這麼做不僅緩不濟急，真的要收集資料也是困難重重。所以追求百分之百的程序與結果再現，並不是能增進科學知識的有效途徑。

Nosek 教授主張真正的 replication 應該是提供檢測原始研究的結論，有證據力的研究結果。嚴謹的證據力不只能確證原始研究的結論，即使不符合也能指出改進的關鍵。換句話說，**有科學意義的 replication，必須能篩選出經得起考驗的假設，也能排除錯不得的假設。**Nosek 教授認為必須研究符合兩個條件，才能成為真正的 replication：其一是與原始研究結果一致能增進論點的說服力；其二是與原始研究結果不一致將減低論點的說服力。要達成兩個條件，一開始就要詳細地探討根據原始研究結論做出的預測，在什麼狀況最有可能成功？在什麼狀況最有可能失敗？完成收集和分析之後，根據事前列出的狀況，評價 replication 的資訊。累積可靠的 replication 資訊，才能讓一個經得起考驗的科學理論走上揭開「知識迷霧」的前進路程。「知識迷霧」是指尚未被 Replication 檢驗的部分，這個概念來自科學哲學家拉卡托斯提出的科學理論發展

觀點。一個有原創性的理論是在一個所有預測尚待檢驗的知識領域，劃出可供 replication 檢驗的一點預測空間，除了能進一步預測的普及化研究(generaliizability)，因為設備條件或測量方法等尚未成熟，無法檢驗的預測，就構成這個領域的知識迷霧。具有拓展知識價值的理論，每經歷一次 replication 與普及化研究的考驗，就會減少知識迷霧，增加 replication 的空間，進入拉卡托斯稱呼的「進化的科學研究綱領」。反之，經不起考驗的理論，發展到某個程度卻出現不一致的 replication，無法調整就會看到知識迷霧擴大的局面，成為「退化的科學研究綱領」。

Nosek 教授主張的 Replication 研究構成成份有研究對象的單元：像是人類參與者；研究操作的變項；收集資料的限制條件；以及預測的研究結果等四種。都是屬於研究的軟硬體成分。而陳瑞麟老師在「認知與評價」談論的 Replication，包含了「理論（背景觀念）解釋」、「實驗模型」、「物質落實」、「實驗結果」四個部分，陳老師統整為「實驗行為結構」。如此主張是為了有效區分三種程度的 Replication：固定理論解釋下的 Replication、物質落實的 Replication，以及實驗結果的 Replication。「認知與評價」一書討論的案例是 19 世紀物理學，檢驗陰極射線的本質是乙太波還是帶電粒子的實驗，主張乙太波假說的赫茲首先設計電荷偵測實驗並獲得正面證據，之後培林與湯姆生分別延用赫茲的一部分設計，卻得到相反的結果。荷蘭科學哲學家瑞德為了討論不同科學家的實驗檢驗理論的功能，提出三種 Replication 程度的評估方法。陳老師認為瑞德的評估方法背後並未區分理論的描述與理論對實驗結果的解釋，因此將科學家的想法區分為背景觀念與實驗模型。

背景觀念是科學家自身對當前課題的問題意識，實驗模型是針對特定問題的實驗計畫，包括目的、手段與預期三個層次。以陰極射線的電荷偵測實驗來說，培林與湯姆生的目的與手段都與赫茲相似，卻因為收集資料的方式不同，導致不同的預期及結果的解釋。回頭看 Nosek 教授主張中的研究成份，似乎是指陳老師的實驗行為結構所指的「物質落實」與「實驗結果」，其實是對應實驗模型的三個層次，每個層次更細分為七個步驟，詳實比較不同科學家的實驗計畫相似程度。我與陳老師的對談中提到還可以加上「分析」這個步驟，就能分析社會科學研究的 Replication 程度。

在對談的後半部，我請教了陳老師有關直接再現(Direct Replication)與概念再現(Conceptual replication)的看法，他認為這相當於「在相同背景觀念下」去企圖使用不同的實驗工具去做出「相同預測結果」的 Replication。但是綜觀科學史，在背景觀念、實驗工具、與實驗結果三個層次的差別，可以組合成至少八種 Replication 樣態，因此他對於心理科學家如此區分 Replication 的看法有所保留。Nosek 教授的評論則指出許多心理學的概念性再現並未明確指出理論的可能缺點，正面實驗結果雖然擴大普及化的範圍，但這些實驗結果並未再做檢驗，一旦通不過嚴格的 Replication 檢驗，就會退化為知識迷霧。

因為單元時間長度的設定，這一集我只做個大致的介紹，讓聽眾朋友了解台灣學術社群也能對講究開放透明，且要求經得起 **Replication** 考驗的科學實踐，做出回應與行動。我很希望能和對 **Replication** 課題有見解的朋友對談討論，提高台灣學術界與社會大眾對 **Replication** 的認識及重視。兩件文本完整內容的網址，都放在節目介紹之中。有興趣的朋友可以到節目的臉書留言或私訊。如果你對本集內容有任何疑問或建議，歡迎到開放咖啡角臉書專頁留言。覺得我們的節目有意思，請分享我們的節目與臉書專頁。我們下一集見。

Show note:

推廣愛好科學的人們認識並且討論 **Replication** 是開始這個 podcast 的主要原因。前陣子 3/27 日推動心理科學開放運動的指標人物 Brian Nosek 在 *PLOS Biology* 發表 "What is replication?" 評論文章，剛好今年初我與中正大學哲學系陳瑞麟老師在 google 雲端文件討論他的著作「認知與評價」，提到的科學哲學家對於 replication 的看法。Nosek 的評論與陳老師和我的對談有很高的互文性，因為陳老師不方便進行即席對談，所以這一集我融合兩份文本，分享經過對談及閱讀後，我對 **Replication** 的看法，以及在中文語境的溝通建議。

與陳瑞麟老師的完整筆談紀錄請見本集朗讀稿。

本集朗讀稿：

本集錄音檔：

Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLOS Biology*, 18(3), e3000691. <https://doi.org/10.1371/journal.pbio.3000691>

施打卡介苗的國家新冠肺炎死亡率較低的說法來源：
<https://www.jsatonotes.com/2020/03/if-i-were-north-american-european-austral.html?m=1>

拉卡托斯的科學研究綱領：南華大學「知識、權力與社會」課程講義
http://www.nhu.edu.tw/~sts/class/class_01_4.htm

Credit: 本集封面圖取自英國杜倫大學哲學系教授 Thomas Rosseter 發表於 2019 年美國地質學會研討論的簡報。下載網頁：<http://www.mantleplumes.org/Lakatos.html>

筆談紀錄：

主持人：陳紹慶

來賓：陳瑞麟

1. 開場：主持人介紹來賓，說明邀請理由：[認知與評價讀後心得](#)

(紹慶)大家好，我是本集主持人陳紹慶。這一集是開放咖啡角的第一次對談企畫，很高興邀請到中正大學哲學系教授陳瑞麟老師蒞臨我們的節目。這次企畫主題「科學研究的 replication」，在第一集上架前就決定要做，也考慮好幾位對談人選。今年一開始科學界有個新聞，2018 年諾貝爾化學獎得主之一法蘭西絲·阿諾德（Frances Arnold），主動撤回 2019 年發表在 Science 的論文，原因是這篇論文的研究結果無法重現。經過[國內媒體報導](#)後，陳老師在個人臉書發表[他的看法](#)，並且提到他在 2013 年出版的專書「[認知與評價](#)」，討論過科學實驗要如何複製，還有怎麼算成功的複製等問題。我趁著開始放寒假的時候找來拜讀，發現陳老師提出的問題，還有他認為的判斷原則，都有點中 2010 年代許多呼籲重視心理科學再現危機的學者，主張心理學家應該認識並做修正的問題。所以我就邀請陳老師擔任第一次對談的來賓。

由於陳老師本人有聽力障礙，不方便進行即時口語溝通，我們先做線上筆記對談，事後再由我們兩人分別朗讀錄音，再剪輯成這一集，所以這一集聽眾朋友可以取得完整的文字對談紀錄。接著就先請陳老師自我介紹，並現身說法他對於實驗複製的看法。

2. 來賓自我介紹、聊聊「認知與評價」的寫作背景；有關實驗複製的看法

(瑞麟)我是陳瑞麟，非常感謝紹慶老師製作這樣一個專輯，並邀請我來當對話者，我感到十分榮幸。我是一位科學哲學研究者，出版過幾本科學哲學的專書和教科書。容我先老王賣瓜，介紹自己幾本相關著作一下。我在 2010 年出版《科學哲學：理論與歷史》（群學），這是一本介紹二十世紀科學哲學發展歷史與各種理論學派的著作；2014 年出版《科學哲學：假設的推理》（五南）是一本討論「科學推理」或「科學方法」的著作；中間 2012 年在台大出版社出版的《[認知與評價：科學理論與實驗的動力學](#)》則是一本科學哲學專著，可說是我建立自己科哲理論體系的一個企圖。

《[認知與評價](#)》這本書的副標題表明我企圖提出一個科學發展的動力學，以便回答「科學發展和變遷是出於什麼樣的因素？」主標題則是針對這個問題的關鍵答案：「[認知因素](#)」和「[評價標準](#)」。這個答案似乎是理所當然。不過，強調這兩個關鍵詞是要對抗國際上的「科技與社會」（STS）和「科學史」的主流觀點，像「科學知識的社會學」（SSK）這種立場，它宣稱科學發展和變遷的主要動力是社會因素，特別是利益和權力等，這些在科學哲學的討論傳統中，被視為科學的「外部因素」，而「[認知](#)」與「[評價](#)」則屬於「科學的內部因素」。「[認知](#)」也就是「[認識](#)」，它包含科學家的認知能力、[認知工具](#)（概念、模型、理論、物質工具等）、[認知方法](#)（傳統

上所謂的科學方法、科學推理)、認知成果(科學知識)等；「**評價**」是使用一些價值標準如「**準確**」、「**簡潔**」、「**一致**」、「**可否證性**」、「**可複製性**」或「**可再現性**」等等來判斷知識能力、工具、方法、成果的好壞，這些標準也被稱為「**認知價值**」。科學的發展與變遷，就是因為不同的科學家以不同的認知價值來判斷各種認知因素，導致理論與實驗的變動和發展。因此，考察一個時代的科學界最重視的「**認知評價**」因素，可以讓我們預測、引導、或扭轉科學未來發展的方向。從這裡可以初步看到我為什麼要討論「**實驗的複製**」。

我先簡介到這裡，看看紹慶老師有沒有什麼問題要問的，再來介紹我討論「**實驗複製**」的相關背景。

二月一日：

3. 主持人談談心理科學的 **Replication Crisis**；為何許多 **replication project** 重視 **direct replication** 甚於過去看重的 **conceptual replication**

(紹慶)謝謝陳老師的介紹。陳老師提到的科學發展「**內部因素**」，是這集節目我想與陳老師交流意見的部分。過去十年心理科學再現危機的主張，是建立在好幾項心理學家協同合作的再現研究，結果都不如預期的證據之上。最近澳洲心理學者 **Alex Holcombe** 公開的筆記整理社會科學協同合作的平均再現率，整體來說不到 **50%**。這個筆記的網址我附在這一集的參考資源，陳老師和聽眾朋友可搭配閱讀。

由此我順便來談我閱讀陳老師的《**認知與評價**》第七與第八章的心得，這兩章的主題就是「**實驗複製**」。我自己從這兩章讀到比較不同研究之間的「**實驗模型**」，是陳老師分析與判斷兩個實驗複製程度的基礎。從我自己親身參與一些協同合作研究得到的經驗，在此提出兩個課題與陳老師討論。聽眾朋友可以先找《**認知與評價**》這本書的這兩個章節來看，對於我們的接下來討論會有更深刻的理解。

第一個課題是「**實驗模型**」的結構，陳老師認為科學實驗模型是科學家的實驗行為背後可表述的實驗運作藍圖，實驗模型是根據科學家的背景觀念設計，並落實於實驗儀器與資源的操作，包含科學家做實驗的目的、手段、以及預期。陳老師認為三個層次包括實驗行為的七個步驟，或者書中所說結構：也就是問題 (**problem**)、對象 (**object**)、裝置 (**apparatus**)、操作 (**operation**)、控制 (**control**)、蒐集 (**collection**)、預測 (**predict**)。就我實際研究活動的心得，我認為蒐集和預測之間還要加上分析(**Analysis**)，因為蒐集後的資料要經過分析才能確定可做預測的模型，社會科學乃至重力波及黑洞等前沿物理學課題，都是需要運用統計方法建立並評估最適合的統計模型，才能做出讓人信服的預測。分析過程是這些領域科學活動最耗費科學家精力的活動。稍早提到的協同合作再現研究，判斷原則是再現研究的分析結果與原

始研究是否一致，再現研究結果的再分析顯示，多數原始研究的分析策略讓小樣本資料顯示符合預期的顯著結果，這些再分析帶出心理科學危機爭議中最常被討論的課題：科學家的可疑研究操作(Questionable Research Practices)，一般簡稱 QRP，據我所知現在還被一些學術倫理教程列為必修主題。因此我對第一個課題的建議是，「實驗模型」的結構應該增加「分析」，涵括科學家處理資料及運用統計方法的活動。

我想請教陳老師的第二個課題是實驗複製程度，是指涉直接再現(direct replication)還是概念再現(conceptual replication)。我要先說明什麼是社會科學的直接再現與概念再現：直接再現是後續實驗運用與原始實驗一樣的參與者條件、測量工具、刺激材料、實驗變項的配置與操作、測試相同的假設。概念再現的後續實驗除了測試與原始實驗一樣的假設，其他條件有部分或全部不相同。由於社會科學的是觀察與測量人類的行為，不同人貢獻的實驗資料必然有變異，所以參與者的背景只要相近，其他實驗條件相同，就是完整的直接再現。例如有個在台北的某間大學進行的實驗找該校學生，接著也在高雄的另一間大學按照相同的實驗條件，找校內學生做實驗，高雄的實驗就是台北實驗的直接再現。

剛剛提到概念再現可以是部分或全部實驗條件不同，所以直接再現與概念再現之間不是全或無的二元觀念。想要進一步了解的聽眾朋友，可以從參考資源提供的兩篇文章，得到更多資訊。「認知與評價」的第八章討論理論假設如何被實驗與複製實驗所檢驗，因為陳老師是用形式邏輯論證這個問題，一開始我以為陳老師的論證主體是概念再現，後來我注意到陳老師將科學知識區分為需要指導及不斷練習的技術知識、以及能明示記載的模型知識，論證主體其實是明示性的模型知識，我才恍然大悟陳老師的論證與開放科學重視直接再現的思想不謀而合。因為協作研究講究以預先註冊備份研究假設與實驗條件，參與的科學家要完全按照預先註冊執行實驗。對比「認知與評價」的論證與現實的科學工作，這個課題延伸出許多有意思的問題：像是做為資料來源的參與者背景條件要多相似才能稱為直接再現？如果語言文化是必要的條件，大部分的社會科學研究是否只能適用原始實驗對象的社會？如果執行複製實驗的人員不具備原始實驗人員的技術，或者不夠熟練，無法再現的結果能不能有效更新科學知識？這些問題都是最近這幾年協同合作再現研究完成後，在學術社群中熱烈討論的問題。接下來我再和陳老師就這些問題討論彼此的想法。

4. 自由討論：「認知與評價」的實驗複製接近科學家認識的 **direct replication** 還是 **conceptual replication**

(瑞麟) 謝謝紹慶老師對我的著作的有價值問題和評論。我與紹慶老師在臉書上互相認識幾年了，除了知道他是心理學家之外，也知道他十分積極地在推動新的科學研究模式，讓我瞭解不少當前心理科學的發展趨勢。紹慶老師提出的很多議題都很有意思，非常值得深入討論。不過，在開始進行回應之前，我還是先根據自己的步調，把我的《認知與評價》一書討論的「實驗複製」問題的相關背景作一個簡介。

我對「實驗複製」的討論，是承襲科學哲學的一個傳統議題，這個議題可以回溯到早期的「科學方法」和「歸納法」的討論。科學哲學的經驗論傳統主張「科學方法」是這樣的：提出一個假設，以「全稱述句」表達，然後尋找個例來檢驗該假設。例如，假設「所有烏鴉都是黑的。」如果我們找到一隻黑色的烏鴉，那麼我們就對「所有烏鴉是黑的」做了一個印證，如果我們找到很多烏鴉都是黑的，我們「歸納地高度印證」了該假設。這是很簡單的例子。在實際的科學裡，通常需要實驗。例如，慣性定律「所有物體，如果在沒有任何外力的作用之下，靜者恆靜，動者恆保持等速直線運動。」要驗證這個定律，得作實驗。問題是要怎麼做實驗？以前教科書可能教我們作斜坡實驗。那麼如果我們要「歸納地印證」慣性定律的話，我們是不斷「重複地」（repeatedly）做斜坡實驗嗎？實際的科學好像不是這樣運作的吧？

研究科學史，我們發現歷史上很多印證理論假設的「大科學實驗」，其實都是只做一次，科學家就接受了。當然，這不是說，實驗科學家的實驗技巧很完美，實驗一次到位。事實上，他在一個實驗成功之前，可能有過多次失敗和調整實驗的過程。但這些失敗和調整的過程，科學教科書不會記載。科學史的研究也不多，除非那些科學家忠實地記錄了自己從開始到成功的一切過程，科學史家才能作記錄。即使如此，一般而言，當科學家的實驗成功了，科學史只會記載他的成功結果。然後，其他科學家就不會再重做他的實驗了，因為已經接受該實驗成功了。接下來是站在他的成功實驗基礎上去進行更多新的研究，而不是去重複檢驗它。換言之，科學史的發展不是像經驗主義者想像的「歸納地印證」一樣地不斷重複一個相同的實驗。所以，當代實驗科學哲學的先驅者 Ian Hacking 曾說過一句名言：「粗糙地說，沒有人曾經重複一個實驗。」當然，Hacking 不是說科學家都不會「重做」實驗，而是「人們不相信實驗結果且懷疑它時，會再試試。」而且，重做的實驗通常會使用不同的裝備、甚至不同的實驗方法和過程。問題是，使用不同的裝備、方法和過程，可以稱作「複製」原實驗嗎？

「重做」或「複製」一個實驗的動機是檢查原實驗，因為「實驗也可能出錯」。理論假設出錯了，才需要實驗來檢驗，那麼「實驗出錯了」，當然也必須由進一步的實驗來檢驗。問題是，用來檢驗原實驗的新實驗，在什麼意義上才算是「檢驗」它？究竟是要根據相同的理論來設計實驗程序，還是根據不同的理論來設計實驗程序？在實驗的邏輯上，我們想要檢驗一個理論假設，我們當然得根據該理論的內容來設計實驗。但這產生一個麻煩的處境是，不同的科學家提出不相容的理論來說明同一個現象時，一派科學家設計實驗的結果證實了自己的理論；另一派設計實驗的結果也證實他們的對立理論。這種情況是否會使實驗失去了檢驗理論的效果？科學史上頗有這種例子。我在《認知與評價》第七章討論的是十九世紀赫茲、培林和湯姆生的陰極射線實驗，就是這樣一個複雜的例子。對於實驗哲學來說，困難在於如何去分析這樣的歷史實

例。因為書中已做了詳細討論，我就不再於此重複介紹，想進入這個對話的朋友，最好可以先去閱讀我的書。

除了上述科學哲學傳統的背景外，另一個相關的背景是「科技與社會」（STS）的討論。在第八章，我也討論一位 STS 大家 Harry Collins 的相關論述。讀者可以去看。其實，除了第八章之外，第五章也有討論實驗複製的議題，第五章是針對十七世紀科學家波以爾的「空氣泵浦」的實驗複製，我批評科學史家 Shapin and Schaffer (1985), *Leviathan and Air Pump* 中的論述。有興趣的朋友也可以讀一下，特別頁 262-277 部分。STS 對於「實驗複製」或「重做」（reproduction）的說明是引入「利益、修辭、協商」等等社會因素去說明，主張決定實驗有沒有被複製或重做的決定因素是社會調控云云。這套說明方式我自己覺得問題很多。但是核心的議題是：什麼因素讓科學家決定一個實驗是成功的，可以用來證實或否定先前的實驗結果？

所以，要回答上述在兩個學術傳統下提出的關於科學實驗的議題（其實，STS 的討論也是針對科哲傳統的，想提出一套「社會學說明」），我自己就發展出一套分析「實驗的結構」的理論，然後一一去回答上述各種問題。可以看到，我的「實驗的哲學理論」是根據「物理實驗」而建立起來的，是不是可以被應用到其它領域上呢？我自己也很好奇。我先說到這樣。

(紹慶) 謝謝陳老師的補充。我順便再次宣傳一下我已經為開放咖啡角在個人網頁設定專屬頁面，從這一集開始聽眾朋友可以直接從網頁聽每集錄音，還有取得節目裡主持人與來賓提供的資訊。所以聽眾朋友想更了解討論內容的話，可以一邊聽節目，一邊查詢我們提供的資訊。陳老師剛剛最後提到的議題：什麼因素讓科學家決定一個實驗是成功的，可以用來證實或否定先前的實驗結果？也是主張心理科學再現危機的科學家們一直在質問的，所以我先來做個整理，來和陳老師討論。

剛才陳老師提出的看法是從哲學家及 STS 學者，這些像是科學研究第一線之外觀察者的角度來看，所以剛剛我藉著介紹自己的讀書心得，提出我認為社會科學研究者會怎麼回應陳老師「實驗模型」的看法。因為陳老師的「實驗模型」是探討科學研究的內部結構，也呼應稍早我提到的兩個課題：第一個是實驗模型結構應該要加上「分析」，還有根據陳老師的理論推論實驗複製的程度，能不能契合 *direct replication* 的要求。這兩個課題源自過去幾年好幾項協同合作再現研究的最重要結果：只有不到一半能得到與原始研究一致的結果。因為像心理學的社會科學都是依賴統計，判斷研究者關切的變項有沒有出現期望的差異。不論是要什麼方法收集資料，過去幾十年許多社會科學研究者都是運用推論統計有沒有出現小於.05 的水準，判斷結果是不是符合預期。心理學再現研究專案顯示高比例的原始研究報告的顯著結果，其實是用了許多手段的成果。許多運用統計的科學領域已經有很多學者發表論文，討論如何偵測與評估這種很難再現的原始研究。不過治本之法是研究者要完整公開自己的研究資料與方

法，在收集資料前做好預先註冊。這些作法在之後節目會不斷提到，這裡先不做討論。

再回來談我和陳老師共同關心的課題：什麼因素讓科學家決定一個實驗是成功的？以社會科學研究來說，除了假設及收集資料的實驗條件，統計分析方法也是重要成份，結果的判斷也是根據 p 值是不是達到預期的顯著水準。因此我認為陳老師的「實驗的哲學理論」，要拿來評估社會科學研究的再現程度或重製程度，需要做些加強更新。有關「分析」的部分我可以給一些建議，剛剛陳老師提到的 Ian Hackings 在 1960 年代就對於統計方法的運用，提出他認為科學家使用統計方法應該留意的原則，科學家如何運用 Hackings 的警示在 Deborah Mayo 的書“Statistical inference as severe testing”有深入的討論。因為我正在消化這本書，之後有機會會在節目裡介紹。

我找陳老師討論這個課題的一個原因是研究重製程度的課題，在台灣的社會科學界還不是一個被重視的課題，但是我自己與國外學界的接觸經驗，讓我感到學術文化由下而上改變的態勢正在茁壯，我擔憂台灣學術界再不正視討論這個問題，耽誤的是一個學術世代的發展，也是讓我決定策畫製作 podcast 的原因。接著就再請陳老師能就我對他的「實驗的哲學理論」更新建議，談談他的看法。

(瑞麟) 謝謝紹慶老師的進一步意見。我先回應紹慶老師對我的「實驗複製」的觀點提出的第一個問題：增加「分析」到「蒐集」與「預期」之間。我完全同意紹慶老師的建議，特別是對那些需要「資料分析」的實驗。我想，我原來的「實驗行為的結構」的框架中，沒有把「資料分析」納入是因為我以「陰極射線實驗」為範例來思考。在那些實驗是被設計來檢驗（敵對的）理論假設的，實驗資料非常直接地顯示出它對理論的證實或否證，也就是說，如果陰極射線被偏轉，那麼就表示陰極射帶負電，如果沒有，就表示陰極射線不帶電（因此是以太波動）。所以，對那些物理實驗來說，如何蒐集到可行的實驗資料似乎很重要，而且在蒐集階段可能包括了分析。不過，我同意紹慶老師，在廣大的各種實驗中，「資料分析」可能比蒐集更重要。例如統計調查實驗，蒐集蒐集並不困難，比較重要的是實驗資料的分析。所以，很謝謝紹慶老師的建議，把「資料分析」加入會讓實驗行為的結構更完整。

不過，我還是想討論一下，「資料分析」和「資料解釋」有沒有差異？因為在我的實驗行為的巢狀結構圖中，從「問題」到「預測」之間，有一個箭頭，代表「如何以問題來解釋針對資料結果的預測」？這一部分是否會被包含在「資料分析」中？

針對紹慶老師的第二個問題，我發展「實驗行為的結構」，並且提議「複製程度」的概念，是為了回應科哲傳統中，在討論「複製」不同實驗面向的問題。我先說明一下科哲傳統對於「複製」的幾個觀點。這是荷蘭的科學哲學家 Radder 提出的看法，我的書第七章第二節有討論。他先把實驗過程分成「物質落實」和「理論描述或解釋」，因此，一個實驗有三種重做類型：第一是重做「物質落實」（含工具） [註：material realization 或稱實驗落實 experimental realization，指按照實驗計畫步驟執行

的研究活動]，不管理論是什麼；第二是重做出相同實驗結構[註：指實驗行為結構，見以下說明]，使用不同的物質落實過程，第三是「在一個固定理論解釋下」的重做，使用不同的物質落實。我自己覺得這樣的「重做」類型的區分有點問題。因為科學史會面對一種「不同理論的解釋下，改良的物質落實的重做，也做出不同的結果。」（陰極射線實驗就是這樣的）。重點在於「相同或不同的物質落實」要如何判斷？「物質落實」又要如何與「理論解釋」區分？我覺得這類的分類問題很大。所以，我自己的理論是使用「實驗結構」來全面取代「複製類型」的區分，我的「實驗結構」的框架中，把「理論（背景觀念）解釋」、「裝備、操作與控制」（相當於「物質落實」）、「實驗結果」融為一個結構，由這整個結構各項目之間的相似程度來比較「複製程度」。所以，如果兩個實驗，其「理論（背景觀念）」、「裝備、操作與控制」到「實驗結果」通通都一樣，這兩個實驗是完全複製，我認為這種完全複製就相當於紹慶老師談到的「直接再現」。至於紹慶老師講的「概念再現」，在我看來，似乎相當於瑞德的「固定理論解釋下的重做」，在我的架構中，也是相當於「在相同背景觀念下」去企圖使用不同的實驗工具去做出「相同預測結果」的複製。可是，這兩種類型的複製，不太能夠完全涵蓋科學史上多樣化的「實驗複製」，例如，「不同的背景觀念、相同實驗工具，不同實驗結果」或者「相同背景觀念，相同實驗工具，不同實驗結果」、「不同背景觀點、不同實驗工具、相同實驗結果」....如此等等，總共有八種組合。這些組合，讀者可以看我的書第十章，討論許多十七世紀的電學實驗。當然，讀者可能會有疑問：那些算是「複製」嗎？問題是，究竟如何定義「複製」？總是有些是背景觀念一樣（重做？），有些是實驗工具一樣（複製？），有些是實驗結果一樣（再現？）。

這可能是科學家和科學史與哲學家的立場差異：科學家關心「實驗複製」是因為他們想知道「如何去檢驗之前的實驗？以保證實驗結論是可靠的？」也許討論「直接再現」和「概念再現」就夠了？不過，科學史和哲學家關心的是：如何提出一個一般性的「實驗複製理論」，可以涵蓋科學史上的複雜多樣的實驗發展？

我對紹慶老師一開始提出的兩個問題的回應如上。先到這邊，再看看紹慶老師有什麼進一步的意見？

(紹慶) 謝謝陳老師的答覆，希望我的一些意見能幫助更新陳老師的理論架構。剛剛陳老師最後提到的幾點，正好說明科學家與科學哲學家看「科學實驗程序」的差異，不過我認為彼此能互相回饋。像是剛才陳老師進一步說明他的理論裡評價複製程度的依據，由「理論解釋」、「物質落實」、「實驗結果」三個部分構成。從科學家，或者至少像我一樣的心理科學家，需要的是能確認資料收集到判斷分析結果符合假設推論的指南，這樣的指南就包含陳老師所說的「理論解釋」與「物質落實」。更有幫助的指南要加上分析結果的判斷原則，這是稍早我提到的預先註冊應該有的條件，也可以回覆陳老師剛才問到的「資料分析」與「資料解釋」，以最嚴謹的預先註冊條件來看，兩者是不同但有具有邏輯性的條件。

今天和陳老師聊了很多，我們彼此交流了很多看法，希望幫助聽眾朋友了解開放科學的哲學基礎就是幫助科學家回歸做科學研究的初衷。這次和陳老師彼此討論讓我想到我的一位外國朋友，荷蘭心理學者 Daniel Lakens 不時呼籲心理學家應該多讀科學哲學家的書，可以反省自己做科學的方法和心態有沒有偏頗。這次對談我的最大收穫是陳老師過去十幾年的研究心得，和現在開放科學提倡者的許多主張不謀而合，中文世界的科學家也可以從中文文獻找到思想靠山。這集節目最後，我想請陳老師給任何領域的科學家，除了您的著作，會推薦那些科學哲學相關著作，能有助他們反省自己從事的研究 replication ？

(瑞麟) 謝謝紹慶老師的總結。也謝謝紹慶老師鼓勵同仁們接觸科學哲學家的著作。至於介紹相關的科學哲學著作，我想，最簡單的方式就是看我的書第七、八章所討論到的文獻。一個可行的起點是 Ian Hacking (1983), *Representing and Intervening*，這本書的後半部討論實驗，但主要是以物理學為主。至於心理學部分，其實現在也有不少科學哲學家以認知心理學為題材討論相關的科學哲學，稱作「認知科學的哲學」([Alvin Goldman 認知科學的哲學應用](#) or 李建會 [心靈的形式化及其挑戰：認知科學的哲學](#))，但是我接觸不多。否討論到「認知心理學實驗」，我就不清楚了。希望這點資訊有幫助。謝謝。

參考資源

陳瑞麟. (2012). 認知與評價：科學理論與實驗的動力學. 國立臺灣大學出版中心.
[google 電子書](#)

Holcombe, A. O. (2020, January 27). Australian Reproducibility Network materials.
Retrieved from <https://osf.io/xkpg8/>

Diener, E. & Biswas-Diener, R. (2020). The replication crisis in psychology. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology*. Champaign, IL: DEF publishers. Retrieved from <http://noba.to/q4cvydeh>

• Evaluating replications designs compared to original:

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). [A unified framework to quantify the credibility of scientific findings](#). *Advances in Methods and Practices in Psychological Science*, 1, 389-402.

Mayo, D. G. (2018). *Statistical inference as severe testing: How to get beyond the statistics wars*. Retrieved from <https://doi.org/10.1017/9781107286184>