

# 內隱聯結測驗發展與爭議回顧

## 一、研究動機

做 Implicit project 的內隱聯結測驗 (Implicit Association Test, 下稱 IAT) 實驗時, 我覺得這項實驗相當有趣, 測試時也發現原來內隱聯結影響我們反應如此巨大, 更好奇他與內隱測驗與外顯測驗的差別、偏見如何影響理性決策。翻閱一些資料後, 覺得 IAT 測驗是一個很有討論性、對心理學研究貢獻良多的領域。因此想試試看做這項測驗的文獻整理與回顧。

## 二、起源

由 Greenwald 等人開始提出一個疑問: 社會行為是處於有意識的控制下嗎? 越來越多研究顯示社會行為通常以隱含或無意識的方式運作。這代表著過去內省法的測驗或外顯的測驗可能無法完整討論行為過程, 因此主張方法論上要更多地採用間接測驗評估受試者的內隱社會認知 (Implicit Social Cognition)。IAT 即在此背景下在 1988 年發表。IAT 可以了解測量我們對某些概念的內隱態度並預測其行為, 是社會心理學中相當重要的一種測量方式。

## 三、實驗機制

最初的 IAT 評估涉及兩對對比概念 (例如, 黑人—白人和好—壞) 的四種關聯的相對強度。測驗過程如下: 由兩個反應鍵 (response keys) 來連結兩個概念與屬性, 比如黑人/白人、不愉快/愉快 (A 組)。受試者需同時辨認螢幕上出現的字或圖像透過按鍵選擇右上或左上的選項。實驗過程中也會交換配對如換成不愉快、白與愉快、黑 (B 組), 詳細流程參照 Nosek(2005), page167。實驗透過測量兩種的反應時間差來了解受試者對於相關概念的連結程度。舉例來說, 若受試者對白人有偏見或歧視, 他可能會將白與壞的內隱聯結大於黑與壞的連結。那他在測驗時 B 組的反應時間應該會小於 A 組。

## 四、質疑與回應

### (一) 信度與效度

Lane 等人 (2007) 的統整研究顯示 IAT 內部一致性的報告也表明它是普遍可接受的 (內部可靠性平均為 0.79) , 也顯示 IAT 的重測信度範圍從 0.25 到 0.69 , 可靠性的平均值和中間值為 0.50 。高於其他的內隱測驗 (如 Stroop task) 。即便如此, 0.50 的重測信度仍然低於外顯或自我報告的測試。也代表著 IAT 測驗很可能可以通過練習來改變。此外對於雙語使用者, 不同的語言施測也會有不同的結果, 這可能與社會基模有關, 但也代表了此測驗很可能無法反映出一個人真實的偏見。

關於效度的部分, 一些獨立的薈萃分析研究指出, IAT 對許類別的預測效度都不是很好, 而且 IAT 的表現並不比簡單的顯式測量佳 (Oswald et al., 2013) 。Forscher 等人 (2016) 則發現隱性測量的變化並沒有調節顯性測量或行為的變化。隱性測量是可以被改變的, 但這些變化並不一定會轉化為顯性測量或行為的變化。也就是說即便現在業界、學校, 許多大型機構 (如 google、NYPD) , 會希望成員去上降低內隱聯結的課程, 即便成員的內隱偏見連結測出下降, 但也不代表他們的行為會有所改變。

儘管在信度與效度中, IAT 在學界有一番討論、質疑, 但他仍然是一個測量偏見很方便的手段。HCD research 這間機構也給出一些使用 IAT 的建議。第一, 研究者需意識到 IAT 只測量關聯的相對強度。例如, 結果只能告訴我們一個人是否更喜歡 A 而不是 B, 而不是他是否討厭 B。第二, 讓參與者在實際研究之前練習測試以提高可靠性試驗可能是有益的。第三, 將顯式測量與 IAT 一起包括在研究中可能會有所幫助。

## (二) 內隱聯結與審慎決策

有些人會認為 IAT 這種測試連自身都感受不到的內隱聯結有辦法真的影響人的決策思考歷程嗎? 還是大部分的人還是會用理性思考比較重大的決策呢。2013 年, Kubota 研究在做完 IAT 後標記為親白人的受試者在做最後通牒賽局的時候會比較容易接受白人給較低的報酬, 拒絕黑人提供較高的報酬。這是一個蠻重要的實驗, 最後通牒實驗時常在研究受試者如何看待公平大於他所獲得的金錢。這個研究顯示人的內隱聯結也是會影響他在做重大的決定的, 並不像受攻擊的論

點那樣。此研究也提醒這些內隱聯結是會進入社會行為的，再次凸顯歧視對弱勢種族的影響。

### (三) 實證研究

2016 年 Hagiwara 等人研究通過檢查非黑人醫生和黑人患者的種族態度對他們自己的影響以及彼此在醫療互動期間的影響和療程參與，他發現許多醫生是厭惡的種族主義者 (Aversive racists)，代表他們外顯種族偏見水平較低但內隱種族偏見水平較高。因為以醫療人員的職業道德會使他們對外顯歧視反感，但卻仍然保有內隱導致。他們發現這導致醫療人員的行為衝突並提供混合訊息，如不一致的語氣和肢體語言，黑人病患也因此會表現出較消極的反應，進而對建議療程排斥。這篇論文將內隱連結測驗應用到醫病關係與醫院管理的領域，並試圖解釋這個現象。

2010 年 Rooth 試著用 IAT 對於招聘的歧視行為解釋成一種自動或內隱的、意識較低的模式，而非傳統的見解是基於偏好的歧視或統計上的歧視 (taste-based and statistical discrimination)。實驗結果表示，當招聘人員對阿拉伯-穆斯林男性有一個標準差更強的負內隱關聯時，邀請阿拉伯-穆斯林求職者的概率會降低 5 個百分點。這表明自動化流程可能會對雇主的招聘決策產生重大影響，為經濟學中勞動力市場的歧視提供新的見解。他在結論時也提出減少內隱歧視的方法是建立更準確徹底的評估應聘者的方法。他也強調將社會心理學的研究方法與經濟學的應用研究相結合，可以獲得很多好處。因為他的實證研究顯示 IAT 在預測與勞動力市場和經濟相當出色，並認為了解歧視背後的“力量”對於最終能夠創造一個公平的勞動力市場至關重要。

## 五、領域延伸與未來發展

### (一) 在其他心理學領域的應用

心理學家在發展心理學領域應用到兒童的 IAT，希望能測驗出何時兒童會產生偏見。Baron 和 Banaji 在 2006 年將 IAT 分別測驗三組不同白人參與者：幼兒園兒童 (大約 6 歲)、五年級學生 (大約 10 歲) 和大學生 (大約 19 歲)。也特別為對年幼的受試者改用朗讀的方式。

研究人員發現三組之間的 IAT 表現沒有差異，都對白人表現出比黑人更積極的內隱態度，而且程度相似。而對於外顯態度，幼兒園和五年級學生對白人的自我報告態度比黑人更積極，但大學生在對白人與黑人的自我報告態度上沒有表現出任何差異。此研究認為內隱連結可能來自於兒童很小的時候就已經建立同種族比較容易親近的連結。反而是會隨著社會化的過程中意識到不該有的外顯歧視，所以可以看到大學生自我報告中對種族間表示平等。

在臨床心理學領域中，Mosca 等人用 IAT 測驗無法忍受不確定性（intolerance of uncertainty，簡稱 IU），IU 是一種關鍵的跨診斷人格結構，與焦慮症狀密切相關。然而傳統使用自我報告的方式來衡量可能忽略一些內省困難或情緒反應。研究人員給受試者同時使用 IU 的內隱測驗（IU-IAT）和不確定性不容忍量表（IUI）兩個分量表進行評估。結果顯示 IU-IAT 具有足夠的內部一致性，並且與 IUI 的相關性不顯著。透過路徑分析模型，IU-IAT 佔擔憂的 35%，代表 IU-IAT 獨立於 IUI 預測了與事件相關的擔憂。這篇文章提供未來焦慮等症狀臨床診斷與後續治療的協助。此實驗也支持如同認知心理學中所說的雙重歷程理論，內隱連結與外顯偏見或許是分別代表人類思考過程中兩種不同的處理途徑（隱性、自動、無意識的／顯性、受控的、有意識）來影響人的思考、情緒與社會行為。

## （二）未來可能的研究

當然，利用 IAT 的研究者首當之急是會需要更盡力設計實驗程序，讓 IAT 的效度、信度可以提升，比如說結合其他測驗方法、受試前先練習等措施，也或許之後研究會提出更好的測量方式來測驗內隱態度，可能不是用時效反應時間來測量，或許是測驗受試者的信念偏誤等方法，來保證其重測信度不會因為多次施測有所影響。

從上述一些論文我整理出兩條可能發展的路徑。我們看到 IAT 這項測量方式越來越受認可，方便、快速的統一化測驗可以使實驗者測量他們想測驗的概念與屬性，如：身材與愉悅、性別與適合的工作...。因此我認為第一條便是繼續測量出不同文化、不同族群的各式各樣的偏見，可以幫助社會心理學繼續探究人類的社會行為。

我們也看到此測驗幫助偏見更容易跨出心理學的領域，應用到管理學、勞動經濟學上。第二條路徑便是促進跨領域的合作，為原本的理論基礎或模型提供新的想法與不同的碰撞。

## 六、心得

在開始寫這份報告之前，我沒想過會有那麼多的收穫。從天馬行空的想法透過讀許多參考文獻逐漸內斂，再到選定題目開始閱讀挑選過的文獻回顧，試著整理出一份報告。讀文獻的過程中也不斷思考這裡用到心理學的什麼理論、什麼樣的統計方法。這些思考都讓我對心理學有更深的了解與記憶，所得到的收穫遠超過我期初選課時對普通心理學這門課的期待。

## 七、參考文獻

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological science*, 17(1), 53-58.

Forscher, P. S., Lai, C. K., Axt, J., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2016, August 15). A Meta-Analysis of Procedures to Change Implicit Measures.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6), 1464.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review*, 102(1), 4.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of personality and social psychology*, 97(1), 17.

Hagiwara, N., Dovidio, J. F., Eggly, S., & Penner, L. A. (2016). The effects of racial attitudes on affect and engagement in racially discordant medical interactions between non-Black physicians and Black patients. *Group Processes & Intergroup Relations*, 19(4), 509-527.

Kubota, J. T., Li, J., Bar-David, E., Banaji, M. R., & Phelps, E. A. (2013). The price of racial bias: Intergroup negotiations in the ultimatum game. *Psychological science*, 24(12), 2498-2504.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test: IV: What we know (so far) about the method.

Mosca, O., Dentale, F., Lauriola, M., & Leone, L. (2016). Applying the Implicit Association Test to Measure Intolerance of Uncertainty. *Psychological reports*, 119(1), 55–70.  
<https://doi.org/10.1177/0033294116658725>

Nelson, T. D. (Ed.). (2009). *Handbook of prejudice, stereotyping, and discrimination*. Psychology Press, 192-193

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31(2), 166-180.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: a meta-analysis of IAT criterion studies. *Journal of personality and social psychology*, 105(2), 171–192.

Rooth, D. O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, 17(3), 523-534.

#### 八、參考網站

<http://www.hcdi.net/reliability-and-validity-of-implicit-association-test/>

<https://qz.com/1144504/the-world-is-relying-on-a-flawed-psychological-test-to-fight-racism/>

<https://implicit.harvard.edu/implicit/user/jaxt/blogposts/piblogpost010.html>