

無證據存在，就是不存在的證據嗎？ ——分析斯滕格爾的新無神論

余創豪

美國哲學家 and 物理學家斯滕格爾 (Victor J. Stenger) 是新無神論運動的主將，他撰寫了許多書籍和文章，發表過許多演講，目的都是為提倡其中心論點：沒有證據表明一樣東西存在，即可以成為其不存在的證據 (“Absence of evidence becomes evidence of absence.”)，以下筆者將這個概念縮寫為 AE = EA。本文目的是從科學史和統計學的角度來分析斯滕格爾的論證，寫作對象是對護教感興趣的讀者，但有些讀者可能沒有科學史和統計學的背景，因此，筆者嘗試避免作出過多的技術性解釋。

簡單地說，斯滕格爾的論點可歸納如下：當我們沒有證據或其他原因去相信某些東西存在，那麼我們便可以肯定它不存在。例如我們沒有任何證據顯示大腳怪、雪人、尼斯湖水怪存在，我們便不應該相信他們存在。基督教的上帝應該在同一個方式下受到檢查。神的存在是一個科學假說，如果上帝真的創造和掌管宇宙，而且積極干預人類事務，那麼祂必會留下一些痕跡，可以由科學去探測試驗。斯滕格爾 (2007) 的結論是，這假設未能通過科學測試 (Kindle locations 115–17)。

(一) 逆否性論證 (Contrapositive Argument)

在某些情況下， $AE = EA$ ，其中的一個情況是逆否性論證，我們可以利用古典邏輯的換位規則而做出這樣的結論： $C \rightarrow E, \sim E \rightarrow \sim C$ 。如果 C 總是導致 E ，那麼沒有發現預期的 E ，就是沒有 C 的證據。例如，如果外面下雨，那麼街道必定濕滑；如果街道沒有沾濕的話，即是說外面沒有下雨。然而，關於上帝是否存在的辯論卻並不是那麼簡單，我們不能絕對肯定如果上帝是存在的話，我們將得到怎麼樣的 E 。

其中一個反對智慧設計和佩利手錶比喻 (Paley's watch metaphor) 的主要論點是：我們在海灘上拾到一隻手錶，肯定它是精心設計的產物，因為我們從前見過手錶，知道鐘錶的構造，但我們卻不知道上帝是如何創造了這個宇宙。科學哲學家索伯 (E. Sober 2000, 46–57) 指出：如果我們清楚知道智慧設計師以何種方式設計宇宙，那麼我們可以說，世界上出現了我們預期的秩序，所以這個世界很可能是由智慧設計師創造的。然而，有神論者觀察現有世界的現象之後，說這個世界似乎是結構良好，於是認為世界起源於智慧設計師。索伯說這是一個形式邏輯上的謬誤，我們不能從 E 推斷到 C ，因為我們不知道 C 如何設計 E 。

可是，新無神論者卻從 $\sim E$ 推斷到 $\sim C$ ，彷彿他們知道了 C 怎樣運作， E 應該會呈現出甚麼模樣。斯滕格爾 (2007) 說，如果上帝是個智慧設計師，那麼人體就必須結構良好 ($C \rightarrow E$)，因為我們的身體有許多糟糕的設計，所以沒有智慧設計師 ($\sim E \rightarrow \sim C$)。換句話說，人體缺乏良好結構是智慧設計師並不存在的證據。他這樣寫道：

我們在三十歲後骨骼逐漸流失礦物質，使我們容易骨折和患上骨質疏鬆症。我們的肋骨沒有完全包圍和保護大多數內臟，我們的肌肉會萎縮，我們的

腿部靜脈會擴大和扭曲，因而導致靜脈曲張，我們的關節因潤滑劑日益稀薄而磨損，我們的視網膜容易脫落，男性前列腺會增大，因擠壓而阻塞小便。

(Kindle locations 590–92)

斯滕格爾假定智慧設計師不應該讓我們在三十歲以後流失骨鈣，應該使我們的肋骨完全包圍內臟……等，但是，他又怎知道智慧設計師應該怎樣去造人呢？是否四十歲以後才流失骨鈣，就是較好的設計呢？還是五十歲？即使上帝將失去骨鈣的年齡推遲至六十歲，我恐怕斯滕格爾仍然會問：為甚麼不是七十歲？我們怎知道三十歲後迅速老化，不是上帝的意願呢？索伯反對創造論的論點也適用於斯滕格爾，我們不能從 E 推斷到 C，也不能從 $\sim E$ 推斷到 $\sim C$ ，因為我們根本不知道上帝如何設計人類。

(二) 檢測問題 (Problem of Testing)

前面的例子十分簡單，我們可以用肉眼觀察地面是否濕滑，但面對一個複雜的現象，我們便不能用天然的感官去檢測，我們需要精密的探測儀器。如果我們使用先進的儀器而沒有檢測到 E 存在的證據，這樣 $AE = EA$ 是有效的，例如我懷疑自己有腦腫瘤，但醫生使用磁共振成像和其他尖端設備掃描我的大腦之後，卻沒有發現任何腦癌的跡象，醫生便可以信心十足地下這結論：沒有證據顯示有腦腫瘤，就是沒有腦癌的證據。但這推論方法需要許多輔助性假設，例如：1. 搜索範圍是有限的；2. 用於檢測的設備是正確且具有高度準確性的。只要其中一項條件達不到，聲稱 $AE = EA$ 便不能成立。斯滕格爾 (2009, 241) 用了一個很簡單的例子來淡化檢測問題的複雜性，他寫道：

你經常聽到有神論者，甚至一些著名的科學家說，缺乏存在的證據不等同有證據表示它不存在，我表示異議。在超出了合理懷疑的情況下，應該有證據卻沒有找到證據，那麼缺乏它存在的證據就可以證明它不存在。例如，沒有人在洛磯山國家公園見過大象，可以肯定的是，如果有大象在公園漫遊，我們應該有證據顯示公園裏有大象存在：腳印、糞便、踏扁了的草。雖然有可能牠們隱藏起來，但是我們仍然可以在超越合理的懷疑下得出這結論：洛磯山國家公園沒有任何大象。按照這種方式，在應該有明確證據顯示猶太教、基督教、伊斯蘭教的神存在的地方中，我們卻找不到證據，這便使我們再次得出超出了合理懷疑的結論：這樣的神並不存在。

相對而言，洛磯山國家公園的搜索範圍並不算大，衛星和生物學家使用的標準設備當然可以偵測到大象是否存在，但在宇宙中尋找外星文明呢？這個研究項目稱為「外星智慧搜尋計劃」(Search for Extraterrestrial Intelligence, 簡稱 SETI)，1979 年加州大學柏克萊分校率先啟動了 SETI；1992 年，美國政府資助 SETI 計劃。雖然科學家到目前為止還沒有發現任何外星人存在的跡象，但也沒有人能確鑿地宣佈：沒有外星人存在的證據，就是外星人不存在的證據。原因不難明白，首先，這搜索空間(宇宙)實在過於龐大，除非我們遍尋宇宙的每一個角落，否則我們無法做出肯定的結論。其次，我們甚至不知道我們是否使用了正確的設備。許多 SETI 參與者致力於探測無線電信號或其他電磁信號，但我們怎知道外星人會像我們這樣使用無線電呢？我們怎知道他們不是用激光呢？如果他們使用超出了我們所能想像的技術，我們可能永遠不會發現他們的存在 (Davies 2010)。

科學史中不乏 AE 被誤當作 EA 的例子，這是因為科學家在不當的地方使用了不當的檢測方法和儀器。1867 年有人宣稱在南非找到鑽石，但沒有人相信這是真的，因為自第 4 世紀以來，印度一直是唯一的鑽石產地。原始的鑽石埋在地球內部深處，當火山噴發時，鑽石跟熔岩一起被拋出來，因此，最有可能發現鑽石的地方是在死火山附近，但是南非大陸卻沒有火山。1868 年，英國派出最好的礦物學家葛戈里 (James Gregory) 到南非進一步調查，審查過許多岩石樣本後，葛戈里斷定整個南非都沒有鑽石。你可能認為這樣的錯誤是可笑的，因為今天南非確實有很多鑽石礦，但當時葛戈里已經用了最好的科學儀器 (Nigel 1980; Morton 1877)。

SETI 的科學家假設外星人使用無線電信號，葛戈里認為鑽石必須存在於火山附近，以上的例子教導我們這一功課：先設概念主導我們使用甚麼檢測方法、在甚麼地方搜索。神存在與否，比尋找外星人和鑽石遠遠複雜得多，我們怎麼能窮盡所有搜索神的範圍呢？我們怎樣才能知道我們的檢測方法是正確的呢？有趣的是，斯滕格爾反對人擇原理 (anthropic principle) 時，他似乎也意識到我們檢測方法的限制。根據人擇原理，宇宙似乎是經過微調 (fine-tuned) 的，這宇宙允許以碳為基礎的生命形式存在，因此它意味著宇宙有一個設計師。斯滕格爾 (2007) 反駁說：

我們期望宇宙中的所有生命都是以碳為基礎的，或者至少是基於重元素化學，但這不就等於說在每一個可以想像的宇宙中也是如此。即使在我們的宇宙中所有生命形式都有相同的基本結構，這並不意味著在其他物理規律和常數下生命是不可能存在的。
(Kindle locations 1437-40)

斯滕格爾在其近作中 (2011) 進一步說：

支持微調的有神論者假設任何生命形式必然與我們非常相似，因此，我們需要氧氣的水，這是全宇宙都需要的溶劑。不過，雖然這適用於我們的生命形式，我們的生命形式卻並不是一個很大的樣本，可能有一天其他生命形式會在宇宙中其他地方被發現。……我敢打賭，遙遠而跟地球沒有關連的外星生命不會基於左旋DNA。那麼，為何我們要規定在每一個宇宙中都有同一樣的生命形式呢？在任何條件充分的系統中，在任何足夠長的時間內，我期待生命能夠產生，我們不要約束生物學。(Kindle locations 2207–21)

到目前為止，我們還未發現有不以碳為基礎的生物存在，但斯滕格爾似乎斷言，缺乏非碳生物存在的證據不能算是非碳生命不存在的證據，因為：一、搜索空間非常巨大，非碳生命可能存在於宇宙的另一邊；二、我們不能假設在其他星系的生命形式和我們一樣。總之，AE 不是 EA。請問，同樣的原則不是也應適用於尋找神的證據嗎？

(三) 貝葉斯概率 (Bayesian Probability)

斯滕格爾說，因為缺乏證據證明上帝存在，神存在的概率是非常小的，這說法是回應英國物理學家昂溫 (S. Unwin 2003) 的有神論。昂溫使用貝葉斯概率，計算出神存在的概率是 0.67。貝葉斯概率是由 18 世紀數學家 and 神學家貝葉斯 (Thomas Bayes) 發明的。為了評估一個假設的概率，貝葉斯指出統計學家首先釐定一些先驗概率 (prior probability)，得到新的相關數據之後，統計學家便更新概率，這變成了事後

概率 (posterior probability)。貝葉斯概率是主觀的，換句話說，這概率量度信念的強度，與頻率進路的概率 (frequency approach to probability) 恰好相反。在頻率方法中，概率是經驗性和客觀性的，例如，如果你想知道擲硬幣獲得字那一面的概率為多少，你可以簡單地擲硬幣很多次，然後觀察結果。然而，在世界上許多事件是不可重複的，也不允許我們進行實驗，神是否存在就是一個好例子。

昂溫估計神存在的先驗概率為 0.5，其理由是：如果我們沒有任何背景資料，那麼我們可以假設每邊有百分之五十的機會。例如，如果我們沒有任何有關奧巴馬和共和黨總統候選人的資料，那麼我們可以假設，其中一人會贏得 2012 年大選而成為美國總統的概率為 0.5。然後，昂溫把支持和反對神存在的證據都放到貝葉斯公式中，再計算六項證據的分量，為每項證據打一個分數，稱之為神聖指標 (divine indicator，簡稱 D)，在每一次考慮到 D 之後，他重新計算後驗概率：

1. 人類善良的證據，如利他主義： $D = 10 \Rightarrow P = 0.91$ 。
2. 人類道德淪喪的證據： $D = 0.5 \Rightarrow P = 0.83$ 。
3. 自然災害的證據： $D = 0.1 \Rightarrow P = 0.33$ 。
4. 內部奇蹟的證據 (祈禱蒙應允)： $D = 2 \Rightarrow P = 0.5$ 。
5. 超自然奇蹟的證據 (神直接干預)： $D = 1 \Rightarrow P = 0.5$ 。
6. 宗教經驗的證據 (敬畏的感覺)： $D = 2 \Rightarrow P = 0.67$ 。

斯滕格爾 (2011) 指出，美國塔夫茨大學 (Tufts University) 物理學家福特 (Larry Ford) 審查過昂溫的計算，並使用相同的公式而提出了自己的估計，福特的結果是 10^{-17} (Kindle locations 2948–50)。斯滕格爾認為福特的結果更為合理。首先，斯滕格爾用他的著名論點 $AE = EA$ 去反對神存在的先驗概率為 0.5：「若我們缺乏任何證據或其他理由去相信

一些東西存在，如大腳怪或尼斯湖水怪，這就意味它是極可能不存在，因此，神的先驗概率應該為一百萬分之一，甚至更小，即 $P = 10^{-6}$ 。」(Kindle locations 2970–72)

按照 $AE = EA$ ，斯滕格爾進一步認為，如果上帝真的存在，祂應該製造奇蹟，斯滕格爾指出世上並沒有奇蹟，因此 $D < 1$ 。他批評昂溫顯示了典型的有神謬論：良善只能來自神，因此昂溫錯誤地定立一個過高的神聖指標， $D = 10$ ，這 D 應該調整到 0.1。福特則指出，人類道德的淪喪和自然災害不利於有神論，他對這兩項的估量分別為 $D = 0.01$ 、 $D = 0.001$ ，斯滕格爾認為這樣的估量更為合理。總之，福特和斯滕格爾的分數遠遠低於昂溫的，無怪乎其後驗概率非常小 (Kindle locations 2974–83)。

如前所述，貝葉斯概率是一種測量信念強度的主觀概率，不同的人有不同的估計是不足為奇的，然而，令人吃驚的是，斯滕格爾這個科學家並沒理會科學方法所應該通過的各項步驟，草率地跳進結論中。當數字涉及主觀評價時，通常統計學家會計算「施測者間信度」(inter-rater reliability)，在這次辯論中，很明顯施測者間信度很低，幸好貝葉斯學派已經發展出不同的解決方案。在這情況下，應該有一個評審委員會，其成員應包括無神論者、不可知論者，以及不同宗教背景的人士。假設各評審員給予不同的分數，可能的解決方案是：

(一) 去頭尾平均值 (trimmed mean) —— 刪除極端的分數，計算餘下數字的平均值。以奧運會為例，一些比賽項目沒有明顯的輸贏標準，如跳水和體操，故此奧運會通常安排七名裁判負責評分，不用說，每位裁判的評估都可能有一些偏差，亞洲裁判可能給亞洲選手較高的分數，歐洲裁判可能會偏袒歐洲運動員，為了減少偏差，確保裁判結果盡量公平，最高和最低的分數都被丟棄，只用剩下的五個分數來計

算平均值。

(二) 平均模式 (model averaging) —— 在這個方法中，研究員混合不同人的信念，由此創建一個先驗分佈。例如，昂溫估計的先驗概率為 0.5，其他四個人的估計可能為 0.4、0.3、0.2、0.1，統計學家可以構建一個混合的分佈線。一些統計學家 (Fellie and Hazen 1999) 警告說，許多數據分析員只建立一個單一的模式，卻忽略了模式的不確定性，平均模式可克服這不確定性。此外，如前所述，頻率方法的概率強調客觀性和經驗性，它無法在不同模式互相衝突的情況下提供任何解決方法，但貝葉斯學派則可以 (Montgomery and Nyhan 2010)。

(三) 敏感性分析 (sensitivity analysis) —— 這是處理不同估計得到不同結果這問題的標準方法，其步驟是：在每一次的分析使用不同的先驗概率分佈，然後比較不同的後驗概率分佈。敏感分析旨在回答下列問題：研究者對結果有多少信心？如果輸入不同的數據，結果改變了多少？假如結果有重大分別，很顯然，概率的估計對輸入數據十分敏感，研究人員必須調查差異的原因 (Felli and Hazen 1999; Skene, Shaw, and Lee, 1986)。

總言之，斯滕格爾使用貝葉斯概率的方式是有問題的，因為他沒有使用至少一個以上的方法來解決不同概率估計的衝突。再一次， $AE = EA$ 是不成立的。

(四) 綜合分析 (Meta-analysis)

貝葉斯概率是斯滕格爾不正確地應用統計程序的眾多例子之一，他還以許多其他統計方法作為根據來支持 $AE = EA$ 。斯滕格爾反對使用綜合分析去衡量超自然現象或上帝存在的證據。對同一題目，不同的研究可能產生相同或不同的結果，綜合分析薈萃這些研究結果，從中獲得一幅整體的圖

像，合併後的指數被稱為規模指數 (effect size, 參 Glass 1976; Glass, McGraw, and Smith 1981; Hunter and Schmidt 1990)。但斯滕格爾 (2007) 說：

這過程 (綜合分析) 是很成問題的，我可不知道在所有科學研究中，有甚麼偉大發現是基於綜合分析而來的。如果多個獨立的實驗均沒有發現某現象的明顯證據，我們當然不能指望一個對組合數據的純粹數學操控，會突然產生一個重大發現。毫無疑問，靈異心理學家和他們的支持者會對我的結論表示異議，但他們不能否認這事實：儘管經過一百五十年的努力，他們並未能提出任何顯示這種現象確實存在的證據，沒有足以引起主流科學界注意的有力論證。故此我們可以安心地作出這樣一個結論：經過這一切努力之後，這種現象很可能並不存在。
(Kindle locations 824–30)

斯滕格爾的要點是，如果一些研究結論得出 $\sim X$ ，我們不應該期望把這些研究合併在一起後會顯示 X ，否則這是人為操控。英國統計學家包達維 (David Bartholomew) 就此作出回應：他不知道斯滕格爾所說的「數學操控」和「突然」是甚麼意思，但他可以肯定綜合分析並不比其他統計程序有更多的所謂「數學操縱」，這做法也沒有甚麼「突然」可言 (Bartholomew 2011, 126)。

斯滕格爾似乎並不了解綜合分析的價值，事實上，綜合分析是糾正許多棘手的統計學問題的標準方法之一，它不僅可用於合成過去的研究結果，也適用於新的研究。例如，貝克 (R. Baker) 和杜爾雅 (F. Dwyer) 進行了八項關於教學視象化的研究，樣本的總人數為二千，但樣本太大是一個問題，

如果所有樣本均放入同一分析之中，任何本來是微不足道的效果可能會被誤認為顯著的效果。有鑑於此，貝克和杜爾雅單獨計算每個研究的規模指數，然後匯集八個研究結果，從而得出一個整體的推論（Baker and Dwyer 2000）。

此外，若研究者把一切數據放在一個測試中，他可能忽略了辛普森悖論（Simpson's paradox）帶來的問題。在辛普森悖論中，基於總匯數據而得出的結論和分別數據的結論可能完全相反。例如，一所大學曾進行研究，探討收生程序是否有性別歧視的傾向。研究員分析了工商管理碩士課程和法律學院的錄取申請人數數據，他們發現，在工商管理碩士課程中，女性錄取率比男性稍高，法學院的數據也出現同樣情況。有趣的是，當兩組數據匯集一起時，研究員卻發現女性的錄取率比男性較低！另一個例子：英國曾進行一項為期 20 年的跟踪研究，追查吸煙者和非吸煙者的生存率和死亡率，結果發現吸煙者的死亡率只有 24%，非吸煙者的死亡率卻高達 31%。然而，當研究員按年齡組別將數據拆開來分析時，竟然得出相反的結論，因為有更多老年人是在非吸煙組而不是吸煙組（Appleton and French 1996）。鑑於辛普森悖論，奧勤（I. Olkin 2000）建議研究者採用綜合分析，而不是將所有數據集中在一次分析中。總言之，使用所有數據的單一分析和綜合分析可以產生不同的結論，這是非常普遍的現象，並不是由於研究員突然操縱數據所致。斯滕格爾不能從單一分析的 AE 必然地推斷出 EA。

斯滕格爾反對綜合分析，開始於他反對超自然心理現象的研究，如超感官知覺（ESP）。早在 2002 年他已經說：

雖然許多超自然心理學家承認仍然無法確鑿證明超感官知覺存在，少數人卻堅持現在有利的證據比比皆是。由於他們無法運用傳統的科學標準去支持這

一結論，惟有自創其他條件，也就是可疑 (dubious) 的綜合分析。這種分析將沒有顯著統計學效果的研究綜合起來，好像它們是一個單一的控制實驗。
(Stenger 2002)

斯滕格爾指出了綜合分析的缺點：檔案櫃效應 (file-drawer effect)，即只有正面的結果會得到報道，負面的往往都被留在檔案櫃裏，乏人問津。後來斯滕格爾 (2008) 及其支持者將對超自然心理現象研究的批評，延伸到祈禱效應的研究。

斯滕格爾把綜合分析批判為「可疑」，是因為他下了一個錯誤的定義，以為綜合分析結合許多沒有顯著統計學效果的實驗。葛嘉時 (Gene Glass) 是綜合分析的發明人，他也是審核本文作者之博士論文的委員會成員之一。根據葛嘉時 (1976)，綜合分析是一種大規模地集合前人研究成果的統計學方法，他從未提過只結合沒有顯著統計學效果的研究。不幸的是，這種對綜合分析的誤解在無神論者的圈子是非常普遍的，例如在《懷疑論者詞典》(The Skeptic's Dictionary) 的網站中，綜合分析被定義為「一種綜合幾項沒有顯著統計學效果研究的數據分析。它把不同的分析混為一談，好似它是一項大型研究的結果」(Carroll 2010)。

2006年，一位名叫湯普森 (David Thompson) 的博士生寫了一篇題目為〈祈禱療效的綜合分析〉的論文，其中審查了23項調查，得出的結論是：祈禱可能導致身體和心理疾病的顯著改善。2007年，美國亞利桑那州立大學教授霍奇 (David Hodge) 也採用綜合分析，檢討了17項研究，得出的結論是：代禱對病人有一些微小的正面影響。綜合分析得出甚麼樣的結論，在很大程度上取決於分析師選擇甚麼文章，綜合分析師完全知道這問題之存在。鄧高達 (J. DeCoster 2004) 表示，綜合分析不奢求真正的客觀性，它只提供了共同主觀性，綜

合分析的作者有時必須作出判斷，然而，這些決定準則應該公開，讓其他學者可以批評。湯普森在他的論文中清楚地說明他的篩選標準，他排除了所有設計不當的研究，最後從 150 篇文章中篩選出 23 篇。霍奇也說明了其選擇標準，其中一個入選標準非常嚴格：他採用雙盲隨機對照試驗（double-blind randomized control trial），在這試驗設計中，參與者被隨機分配到對照組或實驗組，無論參加者和實驗者都不知道誰得到代禱。霍奇選擇的研究都符合美國心理學協會確認療效的標準。因此，說綜合分析「可疑」，絕對是一個錯誤的指控。

（五）數據挖掘（Data Mining）

斯滕格爾拒絕綜合分析，因為在他看來，假設檢驗（hypothesis testing or significance testing）是唯一嚴謹的科學方法。除了否定綜合分析外，他還批評數據挖掘。數據挖掘是一個統計學的技術群集，這包括分類樹（classification trees）、神經網絡（neural networks）、K 均值聚類（K-mean clustering）等，這方法已經廣泛應用於商業智能的領域（Han and Kamber 2006）。據賴勞思（D. Larose 2005）所說，數據挖掘具有以下特點：1. 利用自動計算法；2. 利用大量數據；以及 3. 強調探索和模式識別，而不是預先定了假說或模型。

在斯滕格爾（2007）看來，數據挖掘無非是在缺乏證據的情況下窮追不捨而已。他寫道：

在搜集數據之前，被測試的假設必須一開始就清晰地建立起來，而不是在過程之中或在看到數據之後再加以改變。在數據顯示出一些有趣但意想不到的結果之後，根據數據挖掘而改變假設，這是不能接受的。（Kindle locations 158–60）……自然地，他們（有神論者）往往不願接受典型研究的負面結果，於

是研究員用數據挖掘去繼續尋找，直到發現自己所期待的為止。(Kindle locations 164–66)

斯滕格爾(2003)採用譚伊麗(Elisabeth Targ)一個有缺陷的研究作為例子，以之說明數據挖掘沒有科學價值，然而譚伊麗的研究並沒有使用任何標準的數據挖掘技術。斯滕格爾指責這項研究沒有經過盲目審查，但這不是數據挖掘本身的問題，如果譚伊麗使用其他統計技術，但手稿沒有經過盲目審查，我們可以責怪其他統計方法嗎？筆者肯定大部分統計學家都不會同意斯滕格爾對數據挖掘的看法。事實上，數據挖掘已經用於商業智能(business intelligence)許多年(Han and Kamber 2006)。據賴勞思(2005)所說，數據挖掘的目標是從巨大數量的數據中，提取有用的信息和發掘變項中隱藏的關係。數據挖掘不需要一個前設概念、一個具體問題，又或者一個狹窄的假設，而是開放地檢測數據中可能的模式，因此，數據挖掘被看作是探索性數據分析(exploratory data analysis)的延伸(Luan 2002; Yu 2010)。無疑在數據挖掘中可能會產生一個新的假說，但這並不是甚麼「馬後炮」，而是探索性數據分析的特徵(Behrens and Yu 2003)！事實上，數據挖掘已廣泛地應用在學術研究上，而這些研究項目已通過盲目評審(例如 Yu, Jannasch-Pennell, DiGangi, Kim, and Andrews, 2007; Yu, DiGangi, Jannasch-Pennell, and Kaprolet 2008; 2010)。

此外，典型的數據挖掘技術包括交叉驗證(cross-validation，參Larose 2005)，那是一種重複採樣(resampling)的方法(Yu 2003; 2007)。使用交叉驗證之主要目標，是為了避免資料和模式的過度擬合(overfitting)——過度擬合發生於研究者試圖一次過建構一個完全適切數據的模式。為了補救過度擬合的漏洞，在交叉驗證中，研究員將整個數據集劃分成許多子集，第一組數據稱為訓練集，它的功能是提出最初

的模型，之後研究員用另一子集的數據驗證這模型，如果新的數據和模型無法適應對方，就必須修改模型，如此反覆進行這過程，直到所有數據子集和模型互相適切為止。換句話說，數據挖掘有一個內置的機制，以避免錯誤地將結論建基於單一個分析。由此可見，從數據挖掘得出的結論比單一的實驗更值得信賴。也許斯滕格爾誤解了數據挖掘的要素，以致他強烈地堅持傳統的假設檢驗。

（六）假設檢驗（Hypothesis Testing）

斯滕格爾一方面否定綜合分析和數據挖掘，另一方面卻未能說服人同意假設檢驗是科學方法的最高標準。首先，他誤解了假設檢驗。據斯滕格爾（2007）所說，如果概率或 P 值是 0.05，這意味著：

如果在完全相同的情況下反復地進行實驗，平均來說每 20 次會有 1 次產生同樣的效果，或者這是在任何有限數據測量下，正常統計波動都會產生的統計假象（statistical artifact）。但想一想這意味著甚麼：在醫学期刊中平均每 20 個所謂的發現，有 1 個會是統計假象！（Kindle locations 805-6）

對此，包達維（2011, 125）直言不諱地說：

斯滕格爾犯了統計學的基本錯誤。……正確地說，有可能在一個實驗中數據根本沒有任何區別，卻被錯誤地報告為呈現出差異，這誤報的機會率是二十分之一。但是，每次進行測試時，我們都沒有任何理由去假設測試結果將沒有任何差別，事實上我們根本不知道會否有差別，又或者如果有差別的話，

這差別會有多大。我們可以合理地假設往往有一些無法檢測到的小效應，使用這方法而得到錯誤結論的比例是不可知的。

斯滕格爾試圖利用假設檢驗以支持 $AE = EA$ 是難以成功的，假設檢驗的邏輯是：若假設為真，觀測到現有數據的概率有多大？它沒有回答這一問題：鑑於現存數據，假設是正確的概率是甚麼？假設檢驗是關於模式及數據是否可以互相切合，以科學哲學的術語來說，我們只測試「經驗充分性」(empirical adequacy, 參 Van Fraassen 1980)，如果數據不能適切模式，這並不一定就是意味該模型是不正確的。在一篇題為〈缺乏存在的證據不是有證據表示它不存在〉的文章中，奧特曼 (Douglas Altman) 和布蘭德 (Martin Bland) 曾這樣警告醫學研究人員：「在比較治療效果中，當隨機臨床試驗沒有顯示出顯著的差別時，這結果通常被稱為『負面』，而這會誤導人以為試驗結果真的沒有任何差別。」(Altman and Bland 1995, 485)

(七) 結論

根據前面的分析，很顯然，逆否性論證不能在關於上帝存在的辯論中建立 $AE = EA$ 。逆否性論證可以應用在簡單的情況，但是，我們不知道上帝如何創造了宇宙，不知道祂是如何設計人的身體。檢測問題顯示出尋找上帝的搜索範圍過於巨大；宣稱科學已經證明上帝不存在，就好像使用我們的設備來搜索已知的宇宙之後，便宣佈沒有外星人或沒有不以碳為基礎的生命形式。此外，斯滕格爾沒有正確地使用貝葉斯方法去計算神存在的概率，因為他完全忽略了去頭尾平均值、平均模型、敏感性分析等方法，這樣，基於貝葉斯而得出 $AE = EA$ 的結論是無效的。斯滕格爾十分依賴假設檢驗，

他不認為綜合分析和數據挖掘是可以接受的研究方法，然而，他對這些統計學方法的概念是不正確的。他不明白辛普森悖論，不知道綜合分析和單一分析可以產生不同的結果。他也不明白數據挖掘是探索性數據分析的延伸，根據資料而發展出新假設是普遍的做法。更重要的是，他認為假設檢驗是最嚴謹的科學研究方法，但他卻誤解了假設檢驗的意義。將上述所有問題加起來考慮，筆者認為斯滕格爾無法以科學方法建立 $AE = EA$ 。換言之，他並不能證明沒有神。

徵引書目

- Altman, D., and M. Bland. 1995. Statistics notes: Absence of evidence is not evidence of absence. *BioMedical Journal* 311:485.
- Appleton, D. R., and J. M. French. 1996. Ignoring a covariate: An example of Simpson's paradox. *American Statistician* 50:340–41.
- Baker, R., and F. Dwyer. 2000. A meta-analytic assessment of the effects of visualized instruction. Paper presented at the 2000 AECT National Convention, February, Long Beach, CA.
- Bartholomew, D. 2010. Victor Stenger's scientific critique of Christian belief. *Science and Christian Belief* 22:117–31.
- Behrens, J. T., and C. H. Yu. 2003. Exploratory data analysis. In *Research methods in psychology*. Vol. 2 of *Handbook of psychology*, ed. J. A. Schinka and W. F. Velicer, 33–64. New York: John Wiley.
- Carroll, R. T. 2010. Meta-analysis. *The Skeptic's Dictionary*. <http://www.skepdic.com/metaanalysis.html>.
- Davies, P. 2010. *The Eerie silence: Renewing our search for alien intelligence*. Chicago: Houghton Mifflin Harcourt.
- DeCoster, J. 2004. Meta-analysis notes. <http://www.stat-help.com/notes.html>.
- Felli, J., and G. Hazen. 1999. A Bayesian approach to sensitivity analysis. *Health Economics* 8:263–68.
- Glass, G. V. 1976. Primary, secondary, and meta-analysis of research. *Educational Researcher* 5:3–8.
- Glass, G. V., B. McGraw, and M. L. Smith. 1981. *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Han, J., and M. Kamber. 2006. *Data mining: Concepts and techniques*. 2nd ed. Boston: Elsevier.

- Hodge, D. 2007. A Systematic review of the empirical literature on intercessory prayer. *Research on Social Work Practice* 17:174–87.
- Hunter, J. E., and F. L. Schmidt. 1990. *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Larose, D. 2005. *Discovering knowledge in data: An introduction to data mining*. Hoboken, NJ: Wiley-Interscience.
- Luan, J. 2002. Data mining and its applications in higher education. In *Knowledge management: Building a competitive advantage in higher education*, ed. A. Serban and J. Luan, 17–36. San Francisco: Jossey-Bass.
- Montgomery, J. M., and B. Nyhan. 2010. Bayesian model averaging: Theoretical developments and practical applications. *Political Analysis* 18:245–70.
- Morton, W. 1877. South African diamond fields, and a journey to the mines. *Journal of the American Geographical Society of New York* 9:66–83.
- Nigel, B. 1980. *The world's greatest mistakes*. London: Octopus Books.
- Olkin, I. 2000. *Reconcilable differences: Gleaning insight from independent scientific studies*. ASU Phi Beta Kappa Lecturer Program, November, Tempe, Arizona.
- Skene, A. M, J. E. H. Shaw, and T. D. Lee. 1986. Bayesian modeling and sensitivity analysis. *Journal of the Royal Statistical Society*, series D 35:281–88.
- Sober, E. 2000. *Philosophy of biology*. 2nd ed. Boulder, CO: West View.
- Stenger, V. J. 2002. Meta-analysis and the file-drawer effect. *Skeptical Briefs* (December). <http://www.colorado.edu/philosophy/vstenger/Briefs/Meta.html>.
- Stenger, V. J. 2003. *Has science found God?* Amherst, NY: Prometheus Books.
- Stenger, V. J. 2007. *God: The failed hypothesis: How science shows that God does not exist*. Kindle edition. Amherst, NY: Prometheus Books.
- Stenger, V. J. 2009. *Quantum Gods: Creation, chaos, and the search for cosmic consciousness*. Amherst, NY: Prometheus Books.
- Stenger, V. J. 2011. *The fallacy of fine-tuning: Why the universe is not designed for us*. Amherst, NY: Prometheus Books.
- Thompson, D. 2006. *A meta-analysis of the efficacy of prayer*. Doctoral dissertation, Alliant International University, Fresno, CA (UMI number: 3292156).
- Unwin, S. 2003. *The probability of God: A simple calculation that proves the ultimate truth*. New York: Three Rivers.
- Van Fraassen, Bas C. 1980. *The scientific image*. New York: OUP.
- Yu, C. H. 2003. Resampling methods: Concepts, applications, and justification.

- Practical Assessment Research and Evaluation* 8. <http://pareonline.net/getvn.asp?v=8&n=19>.
- Yu, C. H. 2007. Resampling: A conceptual and procedural introduction. In *Best practices in quantitative methods*, ed. J. Osborne, 283–98. Thousand Oaks, CA: Sage.
- Yu, C. H., A. Jannasch-Pennell, S. DiGangi, C. Kim, and S. Andrews. 2007. A data visualization and data mining approach to response and non-response analysis in survey research. *Practical Assessment, Research and Evaluation* 12. <http://pareonline.net/getvn.asp?v=12&n=19>.
- Yu, C. H., S. DiGangi, A. Jannasch-Pennell, and C. Kaprolet. 2008. Profiling students who take online courses using data mining methods. *Online Journal of Distance Learning Administration* 11. <http://www.westga.edu/~distance/ojdla/summer112/yl112.html>.
- Yu, C. H. 2010. Exploratory data analysis in the context of data mining and resampling. *International Journal of Psychological Research* 3:9–22. <http://mvint.usbmed.edu.co:8002/ojs/index.php/web/article/download/455/460>.
- Yu, C. H., S. DiGangi, A. Jannasch-Pennell, and C. Kaprolet. 2010. A data mining approach for identifying predictors of student retention from sophomore to junior year. *Journal of Data Science* 8:307–25. <http://www.jds-online.com/>.