

「數據、謊言與真相」作品讀後心得

曾經，在一個風和日麗的下午，南陽街附近晃悠著，墊腳石書局蘊含著強烈的沁涼感，步入書局內放鬆休憩，等待赴約時間的到來，隨意看看書架上的書，其中一本不起眼的書，封面寫著「真相」的書立即吸引我的興趣，就是這本「數據、謊言與真相」，稍微翻翻看到第一章，電話響起，結束這短暫的探索，如今，又有機會再好好靜下心來看看這本書，發現書中探討的不僅貼近現實，探討的層面也非常廣泛，數據果然在每個層面都泛用。那麼看看書中的架構，先以眾所皆知的大事件，「美國總統選舉」事件揭開大數據的神秘面紗，接著述說大數據的驚人力量，以及貼近生活的統計分析，不免俗的最後要警告讀者，大數據中藏有陷阱，提醒讀者要小心大數據，這樣的架構讓我回想起大學上哲學課的過程，先探究哲學，再研究思想的力量，最後也是得回歸現實面，否則將迷失在思想的框架中，初步閱讀完，接著要來細細品味本作品中，深藏的大數據奧妙之處。

首先，以美國總統選舉破題，說實在我在看本作品前，並不是這麼關心美國總統的選舉，但是不得不說這件事細思極恐。在歐巴馬總統在選戰中勝出，其實就已經造就一股轟動了，我自己的認知是在那樣的歷史中，有一位不同於以往的總統，是相當不容易的，雖然我聽

說過許多美國人都認為自己的國家是文化的大熔爐，甚至引以為傲，所以對於歐巴馬當選不以為意，但是事實上並不像想像中的這麼不以為意，媒體、網路爭相報導，甚至開始揣測陰謀論，就可以得知其實人民還是很在意，還是認為難能可貴，而我對這件事也不感到意外，畢竟人們還是希望自己的國家有一位強而有力的領導，但是這就看出人們的行為其實與口述的狀況不同，「人們對自己撒了謊」，當新聞報導美國出現史上第一位黑人總統，不僅僅是美國本土的新聞台，國際上的也是如此，甚至台灣的媒體也是大肆報導，收視率反映出民眾對這件事的關心程度，而多如牛毛的民眾接受訪問時，通常喜歡裝的自己早就知道這件事即將要發生，為的就是不想落人後，不想矮人一截，不想表現自己看走眼的狀況，承認自己的誤判，只有少如麟角的民眾會呈現真實的一面，而這件事還不單單只是這樣，在歐巴馬卸任後，居然換川普當選總統，讓這件事更加撲朔迷離，兩位風格迥然不同的名人接連選上總統，重要的是兩人對種族主義上的看法更是詫異，相信許多人在歐巴馬選戰中脫穎而出時，以為美國真正成就了文化大熔爐，不過卻在之後選了一位白人主義至上的領導人，這也直接反映了美國根本就沒有破除「歧視」的框架，「美國的驕傲只是成就在民眾的幻想上罷了」，想要許多種族的人們生活在一塊土地上，著實不易，書中直接點出川普的當選跌破許多人的眼鏡，但是同時也指出普羅大

眾皆是活在謊言裡，沒看到大數據帶來的啟示，背後更是體現人類的優越感，在歐巴馬的勝選中，看到美國人自認對其他種族的包容優越感，而後在川普的勝選中，則是白人種族勝利的優越感，也許這不是甚麼新鮮事，也許只是被人們因為生活繁忙拋諸腦後而遺忘在腦海，如今只是趁勢浮上檯面，抑或是被候選人煽動情緒，對比最近發生的「非裔男子被美國白人警察膝蓋壓頸導致窒息死亡事件」，就是大數據中種族歧視最好的驗證，聽到主流的言論大概分為兩派，一派是認為白人警察種族歧視執法過當，另一派則是認為黑人本身就是行為不檢點惹人議論遭致歧視，儼然成為 M 型的數據統計，簡而言之，種族歧視的人數占比不比想像中的低，我們都以為人類文明有重大的進步，越來越民主，越來越尊重他人，然而大數據在明示所有人種族歧視仍是人類需要努力的課題。

有時候，直覺雖然可以幫助自己抉擇人生，但在大數據領域上可不行這麼草率，在體育競技上，各個聯盟、球隊隊史甚至是球員個人都是會利用許多數據紀錄，紀錄這些美好難忘的時刻，尤其是突破前人所保留的可怕數據，成為下一個世代所追求的標竿，從追求目標到被別人追求的目標，數據就是有這種魔力，讓人趨之若鶩，書中提及 NBA 這歷史最為悠久的也是最至高無上的籃球殿堂，以「NBA 球員大多在貧困環境下長大」作為討論，我也是個籃球狂熱份子，書中筆者

所述的想法儼然就是我的想法，我在沒有特別研究的情況下，直覺真的認為 NBA 球員許多都來自窮困的家庭，家中窮困潦倒，可能還是單親扶養，甚至是需要靠政府的補助才得以獲得學習的機會，因此球員們小時候就認定自己的出身，如果能有機會進入職業球賽聯盟，肯定是不會放過的，大家都知道職業聯盟的薪水優渥，然而，現實卻不如我所想，真的深入調查，才了解我也陷入了「倖存者偏差」的迷思，的確有少部分的球員屬於這樣的原生家庭，但也許是因為他們的影響力太過龐大才導致我有了這樣的迷思，如同作者所言，大多數的球員還是屬於中產階級以上的家庭，第一個數據調查是球員的出生地，事實證明較富裕的都市小孩較容易成為 NBA 球員，第二個調查是屬於球隊球探會考量的重點，一位球員的身高，身高越高在職業籃球中越容易有優勢，一般來說，貧窮的人往往都是比較矮的，這就和我的迷思不同，因此，可以得知我初始的想法是錯誤的，而我也跟許多有著相同興趣的朋友討論過，他們也都跟我一樣的看法——「NBA 球員大多來自貧窮家庭」，由此可知很多人跟我有一樣的直覺，陷入迷思的漩渦當中。

再來看看跟我們生活周遭有關的例子，如何讓孩子成功或是成為權威人士一直都是人們兢兢業業努力一輩子所追求的目標，作者以成功的要點來做研討，分析了得來的調查，發現最重要的要件居然是環

境，我相信如果做街訪調查，教育大概是最為眾多人所聯想到的答案，的確教育是調查中佔比相當高的，但是作者也表示教育頂多讓孩子成為中產階級，不至於成功或是成為權威人士，「習慣是有感染力的」，環境之所以成為作者口中最重要的條件，原因大概就是來自這句話，讓孩子早點接觸相對應的環境，在該領域上的成就也就越成功，比如前面的例子，想讓孩子在籃球領域上有所成就，讓他早點晉身籃球競技中，與其他球員互相切磋砥礪，會是個好選項，然而環境這個選項完全不符合人們所認為的教育選項，若是沒有利用大數據，這也許會成為未解之謎，雖然得到這個真相，也許會使人們進步，但同時也反映另一個問題—「階級複製」，環境是相當重要的，我們已經知道了，但是能否改變並不是人人都有辦法選擇的，因此才會導致階級複製，富裕的人會給自己的孩子更多資源，會有更多的機會去貼近理想的環境，貧窮的人可憐選擇的機會都沒有，並不能給予孩子需要的資源與環境，真相已經浮出檯面，人類社會應著手下一步，實踐社會正義。

與我們的生活息息相關的另一件事，犯罪事件下降作探討，一開始先講述先入為主的觀念，恐怖片或是驚悚片是人們犯罪的主要原因，人們在觀看完恐怖片或驚悚片可能會受到電影薰陶提高犯罪的機率，但是觀察到另一個現象，電影院中一樣也有播映恐怖片或是驚悚片，卻不見得會有犯罪事件，抽絲剝繭下才知曉，原來是酒精惹的禍，一

般的電影院不會販售酒精類的飲品給觀眾，但自家觀賞電影則會邊喝酒邊觀看，也就是這個看似不影響實為關鍵的差異，酒精總是使人衝動，這個例子也反映出台灣常見的問題，父母通常都會要求孩子努力念書，反對孩子玩遊戲，尤其是射擊類的遊戲，理由大多都是怕孩子會有樣學樣，未來會有暴力傾向，以此分析應用演繹法類推，玩遊戲也不一定會造成孩子會有暴力行為，直接推翻台灣父母的說詞，當然後續還延伸出孩子玩手術類模擬遊戲，也不見得會成為醫生的說法來反駁，反而還提升了世代對立的嚴重性，也許哪天大數據也可以在溝通領域上應用，解決世代對立的狀況。

再來說說大數據本身的特性，人們想利用大數據絕對是有理由的，大數據以別於以前的數據統計，有四種新型的力量，第一種力量，提供新類型的數據，許多領域的研究者，其實未必會使用或是相信大數據來統計，他們可能更相信自己生涯中建立的傳統調查數據集；第二種力量，提供誠實的數據，我覺得這是大數據中最為重要的力量，如果說人會有意識的說謊來讓自己顯得更優秀更傑出，更何況是下意識直覺性的撒謊，或者人們真的不夠了解自己，這種狀況對於數據蒐集跟統計來說防不勝防，而利用搜尋引擎可以讓人們匿名使用也不會被調查，像是你永遠問不出的同志比例、仇恨言論、性隱私和顧客的腦袋，大數據儼然成為了最棒的數據蒐集專家，猶如數位版的誠實豆沙

包，使我們終於看清人們真正想要什麼和真正做了什麼，而不是人們說自己想要什麼和做什麼；第三種力量，允許我們把焦點放在人口中的小子集，這句話聽起來有點繞口，簡單的說一小部分的人口就足以提供有意義的資訊，數據統計上除了最怕上述所說的數據錯誤以外，就是怕做了許多沒有意義的資訊蒐集，其實這相當符合人性，每個人都希望自己的努力能夠相對應的換到甜美的果實，但是現實總是背道而馳，蒐集毫無意義的資訊也如同努力對應不到成果，努力白費化成泡影，有的只是無盡的後悔和逝去的時間而已；第四種力量，允許我們進行許多因果關係的實驗，參考 GOOGLE 搜尋引擎中的自動完成功能，就可以知道搜尋次數最多的關鍵字，例如書中提及的搜尋虐童及墮胎，以虐待兒童來說，通報的案件並沒有增加，事實上只是隱瞞了真相或是等不到救援而放棄了，另外以墮胎來說，合法醫院登記墮胎的次數並無增加，不過搜尋非法墮胎的次數卻在短時間累積，這其實就是因果關係的驗證，以常理判斷，登記墮胎的次數並不會造就搜尋非法墮胎的次數增加，那可能的狀況就是人們怕自己被政府重罰，所以尋求非法手段墮胎，大數據的四種力量為各個領域開啟跨世代的統計分析，未來顧客導向的作法將會越來越多，致使人們生活更趨便利，但相對的，資訊安全以及個人隱私亦有可能攤在陽光下，被有心人士利用，像是中國雖有讓他們驕傲的天眼系統，但其實天眼很多情況下

是違反民主，違反憲法保障人民的權利。雖然書中只敘述四種力量，但其實還有其他兩種可延伸的應用，其中一種應用是尋找分身，這個延伸的應用挺有趣的，就是把個人的特質興趣當作篩選，可以找出與自己有相同特質或是興趣的分身，分身所反映出的其他面向也許也會跟你一樣，或是分身所喜歡的東西，你有可能不知道，但你其實也是喜歡的，這就讓我想起「世界上有 3 個和自己長得很像的人」，說的大概就是這樣的感覺，有的說是外表很像，而個性很像也是有機會的，順帶一提，GOOGLE 搜尋輸入「世界上有」，第一個出現的就是「世界上有 3 個和自己長得很像的人」，看來這句話也成為另類的大數據的標的呢，搞不好以這句話作為研究基礎也頗有一點意思；第二種延伸是實際的實驗，A/B 測試，書中以標題作為考量，準備 A 及 B 兩種方案，可以驗證大數據的分析結果，這是最不同以往統計分析的能力，這大數據才專有的，先是能夠提供兩種不同的方案，就已經成為淘汰以前統計分析研究方法的理由了，學理分析後可以用實際的實驗做驗證更是寶貴的，畢竟許多學裡的方法，未必能在實際狀況中做出眼見為憑的效果，四大力量再加兩種應用，大數據成為未來重點研究的項目之一。

前面敘述了許多大數據的精彩表現，總是得拉回現實面審視大數據的問題，雖然利用 GOOGLE 搜尋引擎搭配大數據平台，可以解決人

們說謊，也就是資料蒐集錯誤的問題，但是卻無法解決維度上層面上的問題，舉例來說，你無法利用大數據預測哪家的股票賺錢，根本的原因是複雜程度，也就是維度跟層面的問題，股票的變數太多，每個變數牽扯的狀況更多，就好比大學上的微積分以及進階的微分方程式，以一句話形容就是「蝴蝶效應」，牽一髮動全身，也可以說股票本身的變化，就是一個大數據，而人類要追上它變化的速度，統計它的型態，預測它的趨勢，近乎是不太可能，更何況還不能肯定是準確的；另外，我自己也發現大數據另一個問題，如果以作者所述的例子，以GOOGLE 或其他搜尋引擎當作工具，必定會出現許多誤差，搜尋跟實際狀況未必成正比，假使以「搜尋自殺」為例，有勇氣搜尋自殺，可不代表有勇氣自殺，就意味著真正自殺的人並不是像搜尋的一樣多，又或者另一種狀況，未懷孕的女人有可能搜尋「懷孕不能吃什麼」、「懷孕不能做什麼」的情況，但要以搜尋來判別同一時期有多少人懷孕，顯然就不一定準確，也就表示搜尋的項目未必對等數據的目標項目，這兩個大數據的問題，一個是技術上的問題，一個是根本上的問題，老實說我不認為可以克服，因為大數據本身還是得考量到隱私跟道德的問題。

最後的結論，大數據告訴我，很少人看到最後一頁，那我必須說我是那很少人中的一個，在我未製作大數據之前，這個事實我想我也

是認同的，的確很少人會有那個恆心毅力，一開頭就要做到最後，看書也是如此，人生的旅途更是如此。提及大數據，可能會想所有的國家、學者或者公司都致力研究大數據，而我腦海中出現的第一個國家，必定是中國，前一段有提及「大數據本身還是得考量到隱私跟道德的問題」，中國雖然也有許多大數據成功的案例，但是也有許多離經叛道的狀況，像是網路控管、天眼系統、行動化支付，網路控管限制人民上網的自由，只能看到國家肯給你看的，這也就是為什麼中國人民都認為台灣是中國的但台灣人從不這麼認為，說的巨觀一點應該不只是網路控管這麼簡單而已，要說所有的行為都有控管的規範，再來，天眼系統總讓人覺得被監視，政府直接擺明監控所有人民的行蹤，好處大概是犯罪了逃不了，壞處則是做什麼完全沒隱私，經過電腦及大數據平台的整合，照到哪個人，就可以顯示出他個人的所有資料，無所遁形，而且說你是就是，說你有就有，不容質疑資料錯誤的可能性，最後是行動化支付，我知道許多中國人真的以行動化支付為驕傲，把它當作一個進步的指標，甚至當作一個台灣落後的理由，不過也許他們沒料想到的是帳戶內的一切，都是攤在陽光下，這會有什麼問題呢？假設你是個生意人，為了資金周轉把帳戶內的所有錢都匯出到國外的銀行，但可能被解讀成你想要捲款潛逃，如果再加上一點政治或社會動盪不安，更會被放大檢視成叛國也說不定，另外我也想起鴻海董事

長郭台銘先生曾公開在新聞上發表，如果未來每個人都有自己的大數據，要娶他的女兒要先交出自已的健康大數據，那豈不是沒有隱私？這就是為何我認為大數據的問題極難克服的主因，「大數據本身還是得考量到隱私跟道德的問題」，如果研究本身不需考量道德，那麼「分歧者」的電影劇情將會真實上演，以後實驗研究就沒有人道主義可言了。看完本書後，最感到可惜的地方是作者做的都是美國的研究，我也好希望哪天能看到台灣特有的大數據，看看風情不同所帶來的差異，根據這些差異，也許能夠促使我們的社會更加的進步，科技會使人成長，亦可能使人腐化，大數據的研究與應用，在未來將會是各國爭相研究的重點之一，至於是使人成長還是腐化，還是得回歸人類本身如何去使用大數據，如何在科技與人性上取得平衡。