政治科學中形式理論的運用與瓶頸 ── 從賽局理論談起

石 之 瑜*

本文探討數理化模型（即經過嚴格定義的符號）在具體指涉事物、或事物與事物之間的邏輯關係時，所發展出來的知識，是什麼意義的知識，並檢討在使用形式理論時，政治科學分析要注意什麼事。文中具體針對效用、邊際函數、影子價格等觀念有所反省，同時藉由囚徒困境模式及其應用上的限制，說明形式理論在知識論上的一些問題。

關鍵字：形式理論、賽局理論、囚徒困境、影子價格、邊際效用、政治學方法論

* 台灣大學政治學研究所專任教授。
投稿日期：二〇〇二年六月二十日；接受刊登日期：二〇〇三年三月二十六日。
東吳政治學報/2003/第十七期/頁 1-19

一、形式理論的簡化功能

在當代主流政治學期刊上被視爲最先進的政治學研究，當非數理形式理論（formal theory）莫屬，它和統計相關性分析最大的不同，在於統計分析是用歸納的方式蒐集資料，從繁多的資料中整理出相關變數之間系統化的聯繫，並用以爲檢證理論假設，或用以爲建立新理論的啓發。相對於此，形式理論（或模型）是用演繹的方式來分析人的行爲。本文主要的討論焦點，是關於形式理論在知識論方面的意義。

爲什麼要做數理化的政治模型呢？嚴格講，既然數學作爲符號邏輯也是一種語言，則凡是日常語言可表達的，照理都可以轉化（或翻譯）成符號邏輯來表達，比如用 n 代表數量。如果語言是一種結構或邏輯，可以指涉具體的事物或動作，只要事物或動作存在，則人所表達的語言只是一個技術問題，不涉及事物或動作的本質，且表達技術愈好，指涉的事物或動作愈清楚。在事物或動作複雜到日常語言難以清楚呈現時，藉由工具性強，操作性高的符號邏輯來代表，有利於對複雜現象的溝通，以及對事物或動作的控制。

一般的數理形式理論家相信，所有的語言都可以用形式邏輯來表達，所以假如政治科學是一種語言的話，當然關於政治科學的知識也可以用數學來表達。好處是，數學的定義非常清楚，邏輯非常嚴謹，如 1+1=2，如果定義域與對應域之間的函數關係並非 f(x)=x 的關係，則表示 1+1 等於 2 未必適用，即 f(1)+f(1)不等於 f(2)。有時候，某種非線性的函數關係比線性關係更貼切，社會科學裡常用得到的這些非線性函數包括三角、對數、指數、微分、間歇函數等等。這些定義精準的函數，經由數理化模型（即經過嚴格定義的符號）在具體指涉事物、或事物與事物之間的邏輯關係時，所發展出來的知識，是什麼意義的知識？在使用形式理論時，政治科學分析要注意什麼事？

二、目標導向的理性

由於數理符號的特色在於邏輯性強，故以數理模式分析人的行爲，難免假定人的行爲有其邏輯一致性，即使這種一致性必須透過極其複雜的函數才表達得了。既然人的行爲具備內在一致，則分析單位便應當以內在一致的個體爲主，[1]而不試圖分析沒有內在一致性的（或非理性的）分析單位，此爲方法學的個人主義關於人性的假定。故數理模式要研究人的行爲的話，首先要確定分析的對象必須是可以「個人化」的單位。所謂個人化，不見得是個人，也許是國家或其他群體，但是群體之內的成員之間自須具有內在一致性，如此國家或群體才具有個人化的特質。是否具備內在一致性的主要指標，在於個人化的分析單位有無自己的行爲目標。故研究議程上的首先要著，便是找出各個個人化的分析單位有何「行爲目標」，照提倡形式邏輯最力的學者William Riker的說法，此即爲界定行爲者的目標理性（Riker, 1995: 24-25）。一旦目標界定，就等於在邏輯上解釋了行爲。當解釋不了的時候，就表示目標界定得不對，而必須另外界定目標。研究必須一直進行直到找到與行爲一致的目標後，才算是成功解釋了行爲。研究者不會去挑戰目標理性，否則形式理論就失去戰場。

因爲形式理論家認定人的行爲是目的導向的，所以所有形式邏輯的模式，就是要追蹤研究對象如何追求其目的──這種追逐目的的行爲，對經濟學家來講，稱之爲個人效用的極大化。同時，在既定目標的限制之下，可以藉由線性規劃的模式，找出最小成本的行爲選項。如此追求目標的行爲，即是效用極大化的行爲。經濟學家的研究語言雖然比較不具體，因爲他們用效用的概念取代特定的目標，但也之所以經濟學家沒有那麼強的需要，非去把所有目標的先後順序界定，才算完成研究，這種做法的確超越了研究者的能

1. 有關社會科學中對於人性假設的不同主張，可參見 Brodbeck, May (1968).

力。故經濟學家乾脆不管具體目標，只討論效用，這樣每一個人即使各有不同的目標順序，都是效用極大化的行爲者，每一個行爲選擇各有其機會成本、機會利益。由於效用是主觀的，還是要靠一些外在的可以定義的符號爲基礎來衡量，符號有很多種，如貨幣（或股票、債券、信用等），而且分析時不能只看今天，還要看對未來價值的判斷，於是要加上時間變項，因此益加顯得數理模式的威力。

相對於經濟學家的貨幣，政治學家比較熟悉的效用單位則是權力，因爲權力可以用來達成有價值的目標。[2]國際政治學家甚至用權力來定義國家利益，權力就是衡量外交決者的效用單位。權力怎麼衡量呢？過去也許用土地來衡量，也許用GNP來衡量，或用軍隊的數量。複雜的計算更包括士氣、素質、管理體制等等（Cline, 1977）。麻煩在於權力的內容有各種各樣的依據，很難標準化，則也許可以用給分的方式。比如，某國在聯合國獲多少票可以算幾分，某國有多少財政赤字又應當扣幾分等，這樣集思廣益列出很多指標，每個指標有個給分的機制。這種給分機制是靠歸納社會上菁英的意見，因此不能免於武斷。較簡單的方式是，是依照結局來排序，依序給分。比如在兩個行爲選項與兩種可能情境的交叉下，出現四個可能的結局，研究者根據研究對象在各個可能結局中的權力得失（或利得，即payoff），最好的結局給 4 分，最差的給 1 分。[3]

之所以提到這些不同的計算方式，是因爲方式不同，則分析出來的結果也有意義上的不同。比如說給分，如何讓金融赤字和戰略結盟多少等量齊觀？這兩件事是不是南橘北枳？不過由於愈複雜的給分基礎，愈可以把利得細分成好幾個面向，如此涵蓋的範圍較廣，這又是個優點，可是每個面向之間的重要性如何拿捏通常十分武斷，這是缺點。

如果用排序給分，依照可能的結局各給多少分，則每一種結局的差距都

2. 貨幣與權力的類比，參見 Parsons, Talcott (1969).

3. 其他常見給分方式包括序數效用（ordinal utility），或簡單的偏好排序（preference ordering）。

變得一樣，如果最好的和次好的之間差距不大，而次好的與第三好的之間差距很大，給分時分別用 3 分、2 分、1 分，顯然又不大適合。所以在做形式邏輯時，必須事先判斷所進行分析的對象是什麼性質，才能在界定用什麼量化方式來代替具體目標的表達。

形式邏輯中最流行的是賽局理論（game theory）。賽局理論除了最早以研究行爲者的目標或效用最大化策略，來解釋行爲的選擇，之後也開始發現，問題並不只在於自己是否達到目的，而是看別人是否達到目的。如果只在意在自己是否達到目的，則別人的行動選擇對自己比較重要，因爲他們的行動影響自己的結果，至於別人的結局就不那麼重要。但因爲目的被量化了之後，使得每個行爲者可以與其他行爲者比較，誰得的效用多。於是目標發生質變，不只是追求自己具體的目標，而是要比「別人」得的多。以致於每個行爲者的效用，要根據他人得到多少效用來決定。筆者親身經歷的一個當時很受注意的例子，是一九八二年美國灰狗巴士受雇者大罷工，原本大家認爲談判不困難，因爲雙方可接受的待遇調整範圍有重疊，但竟然曠日持久，因爲雙方對峙氣氛濃烈，都不希望對方達到目標，所以即使原本自己經過計算可以接受的結局，都變成不能接受。

這裡儼然帶有馬克思主義者講的相對剝奪，就是不問自己具體達到多少目的，如果自己和別人差距愈多，感覺就愈差，故同時有兩個模式，一個是行爲者具體得到多少分，第二個還包括與別人比較之後，修正過的效用所得。自己比別人差得越多，即使絕對效用值增加，可是總效用值卻更低。這種分析風格後來產生了革命理論中有名的「J-curve」論（Gurr, 1970），即革命發生在大家所得都提高，但是實際所得趕不上預期所得之際，令社會大眾產生一種被剝奪的感覺。另一種看差距的方式，不是看自己比別人差多少，是看自己是不是比別人多，只要自己比別人多，即使絕對效用值降低了，反而還更快樂。這一類的行爲不在一般賽局理論的顧及範圍之內。

至於怎樣給分，通常形式邏輯家不會仔細教我們，因爲這個過程是很難客觀的，等於要求讀者在研究者所定義的利得或效用結構之下進行分析，如

果接受研究者的假設，決定了怎樣給分的方式，剩餘的就是數學能力的問題，照理大家應該就會得出與研究者一樣的結果。如此，形式邏輯得以號稱超越了主觀，達到了科學的境界——這就是形式邏輯的威力。

三、邊際值的計算

形式理論藉由數理化，規定行爲者是理性一致的。如此藉由符號邏輯的威力，帶動社會行爲研究的科學化，只要順著理性主義與個人主義的方法論，政治科學的客觀性便可以提升。但假如對於人性有不同的假定，不論是精神分析家以爲的潛意識、後殖民作家揭露的文化混血、儒家推定的差序格局或女性主義者提倡的陰陽共生，都不主張人性的內在一致，更強調情境與關係，亦即人的存在形式與思維方式都是高度社會化的、透過學習的。對這些不以理性與個人主義爲前提的哲學家，形式理論不過是另外一套哲學主張（徐振國，2002），則用形式理論所研究得來的知識，不如說是用形式理論實踐出來的知識，也就是說，形式理論不是外於知識的分析工具，爲分析家所用而已，更是知識本身的內容，使用形式理論與不使用之間的差別，不僅是方法的區別，更是關於知識內容乃至於宇宙本體假定的差異。

事實上，形式理論從來不僅止於是形式理論而已，而已經成爲公共決策中極重要的模式。[4]換言之，以個人化的立場界定目標，成爲一種決策的規範。不依照這個規範的結果，就是違背了理性主義，或者否定了理性計算賴以進行的個人身份位置。Riker很明白地說，假如不能界定目標，或沒有追求目標的行爲，就沒有理性抉擇，也就沒有形式理論，自然更不用提社會研究的科學化了。

理性的人有主體性，這是行爲者作爲一個單位來計算利益的前提，只要人有理性，數理化分析家根據在效用函數中適當的點上所計算的斜率，便可

4. 例見 Stokey, Edith, and Richard Zeckhauser (1978).

以得出行爲的指導原則。所謂效用函數斜率的變化，是因應社會實際環境的限制所作的調整，代表外在環境限制的函數（通常假定爲線性）與效用函數相切時的斜率，乃是效用最大值之所在，即爲效用最大化，這一點是經濟學裡稱爲邊際效用極大之所在。「邊際」作爲形式理論的關鍵概念，可說是人的理性的極致表現（Howey, 1999）。目標導向的行爲是邏輯化分析的前提，沒有目標導向的行爲，就沒有邏輯化的分析。同時，目標的量化也是必然的要求，否則效用極大化或痛苦極小化的方程式無法寫。邊際效用、邊際成本、邊際收入等概念於焉大派用場。

在公共決策中，最大邊際的計算可以幫忙做決策，其中常用的概念就是所謂影子價格（shadow price）（Little and Scott, 1976），蓋公共政策決策時應當要計算服務對象的總體效用，相關的常見概念還包括所謂的消費者剩餘。影子價格告訴決策者許多訊息，比如人命值多少錢。以政府補助救護車的政策爲例，某一年花多少預算在救護車上，要根據往年救護車能救多少人來決定。這時救護車數量與救回生命的目標之間，根據歷史所建立的函數關係，透露出在不同的預算額之下，每再多救一條人命必須多花多少預算。人命的價值就在這個影子價格的計算中揭露出來，於是決策者可以判斷是不是要花更多（或少）預算在救護車的補助上。同理，在已經算出殘障人士每多跑一公里路，須多花多少預算的歷史函數之下，則補助小福康巴士的預算額，隱含了殘障人士多走一公里的社會價值（或成本）。

當然，歷史函數怎麼計算是關鍵，因爲這個函數決定了投入與目標之間的關係。要計算歷史函數，就必須定義目標，亦即將目標量化。如前所述，這是個極爲主觀的過程，故不同的形式理論家可能因爲定義不同，而得出不同的理性建議。比如，每次一輛救護車的使用救回多少生命，沒有客觀的計算方式，必須有複雜的定義。複雜化常常給人客觀的感覺，但不免仍屬於人爲模擬的計算方式，其實並不能代表更客觀。

如果不是政府在做決策，而是個人，形式理論家似乎理所當然地假定個人是理性的，亦即個別行爲者知道自己的效用函數。雖然在臨床精神醫學中

大量案例顯示，人的行爲受到各種不能用語言表達的力量在影響，故在講精神分析時，或社會心理學中，極少談論理性的概念，而是談如何融入社會所認知的角色秩序。從這些其他學科的角度看，形式理論其實是在建構一個社會共同且唯一的角色秩序，名之曰理性行爲者，教導行爲者如何有意識地建構自己的效用函數，一方面鞏固理性主義作爲規範，另一方面讓行爲者成爲政治科學可以分析的完整對象——這是形式理論在本體論上最具威力的地方。

四、「囚徒困境」與「重複囚徒困境」

以形式理論中最著名的囚徒困境爲例，這個行爲模式的目的，是在說明人透過學習，即使基於自利而在短期內做出不利於自己的決定，但長期裡最終會回歸最理性的方式，修正自己的行爲選擇。最原始的囚徒困境模型是兩個囚犯被隔離訊問，招認就能減低刑期，不招會被罰，並假設只要一人招認，檢察官就有足夠證據能處罰兩個人。當兩人都招時，兩人都會中度受罰，但當兩人都不招，則檢察官只能輕罰兩人。重點在於，如果只有其中一方招認，招的那一方就能獲釋，不招者則得到重度懲罰，每個人都會去想（對方）到底招不招，然後發現，不論對方招不招，我自己不招的好處都大於招認，所以兩人基於理性，就各自發展出主導策略（dominant strategy），即都選擇不招，因此都受到中度懲罰。

但假如這個賽局要重複不斷地進行，當然最理性的方式是大家養成合作的默契，都不招認，因此都只受到輕度懲罰。假如默契不養成，兩人每次都受到中度懲罰，顯然不理性。不過，要是賽局只有一輪，基於不信任，大家就都會訴諸短期理性，寧可受到中度懲罰。這就是行爲者學習模式。

將囚徒困境應用到核武競賽，並分析美蘇制定國防預算，假定每年預算透露各方有否遵守限武協定，若一方違反協定大編預算，另一方當年就會落後。爲了避免被欺騙而落後，美蘇照理都會有強烈的意願來欺騙（違反限武

協定），但是美蘇兩強爲什麼遵守所簽的核子限武協定，甚至在到期之後都自動繼續遵守？原因就在於這是長期不斷進行的賽局，雙方自然會發展出合作的默契。對任何一方而言，最好的結局就是大家都不欺騙，欺騙是對自己不利。囚徒困境論者證明時間是重要的因素，即賽局玩一次和玩無限次的結局是不一樣的，只要美蘇兩大超強的狀況持續下去，軍備競賽不會不可收拾，囚徒困境似乎成爲是最好的解釋性模型（Kaplan, 1957）。

囚徒困境也可以運用於雙首長中的府會之爭。如法國總統與總理的兩者任期不同（總統七年，總理五年），既然雙方都會想，在第五年時必然不會合作，因爲沒有第六年需要考慮，所以第四年時就也不必爲第五年的合作氣氛著想，則第四年就不會合作。依此類推，第三年時就不必爲第四年著想，第二年不會爲第三年想，第一年不會爲第二年想。結果，雙方自第一年起就摧毀了彼此的信任。[5]

道理在於，玩賽局的次數是有限的，n 不是無限大，如果是無限大，就會繼續合作，因爲考慮到 n 乘上兩人合作的利得，遠大於 n 乘上兩人不合作的利得，所以兩個人終究會學習合作。但 n 若有限，雖然合作的利得較大，但是因爲知道最後一次不會合作，以致於倒數第二次也不合作，依此類推，於是一開始就不合作。所以在雙首長制下，照囚徒困境來看，總統與總理如果屬於不同黨派，合作的機率是很小的。甚至屬於同一黨派，假如總理有心問鼎總統寶座，而總統也有心連任，他們的合作一開始就會有問題。

囚徒困境也可以好多人一起參加，就是所謂的 n 人囚徒兩難（Barry and Hardin, 1982）。比如，城市裡有千萬戶人家，各家要不要把垃圾好好裝到袋子裡，還是偷偷亂倒？最好的情形就是社會上大多數人都配合（好好裝），只有自己一家不配合，如此環境也不會被嚴重污染，自己又省事。但假如每個人都這樣想，環境就遭毀掉，每一家都受害。長期裡當然大家最理性的做法，就是好好裝垃圾。不過，當納莉颱風在二〇〇一年過境台北造成災難，

5. 也可以用來分析中共派系政治中首腦與接班人之間的權謀關係，見 Huang, Jing (2000).

有少數家庭將垃圾亂丟，引起每家都丟垃圾，便是信任被破壞，則理性的行爲者可能就估算著，要是在這次不丟，以後沒有機會丟，就占不到便宜的心裡下，使得 n 人囚徒困境的默契遭到破壞，於是每家都受害。

Mancur Olson 用囚徒困境分析早年工人不加入工會的心理，因爲只要其他人都加入，自己即使不加入，資方的讓步仍可利益均沾，這就是 freerider（Olson, 1965）。n 人囚徒困境也用來分析革命，認爲如果只有自己去推翻政府，別人不參加，那自己會被政府抓去處決；最好的情形是所有其他人都革命，自己在家裡坐享暴君被推翻的好處，大家都如此想，於是就沒人上街革命。但爲什麼個革命仍會發生呢？因爲每個人的風險傾向不同，風險傾向愈高者，在革命成功機率不變與利得不變的的條件下，比風險傾向低者，愈容易在革命的較早期，亦即成功機率尚低之際便選擇加入革命。風險傾向中度者，會在若干人加入後隨之加入。風險傾向低的人會在多數人加入後再加入。環環相接，故革命仍有發生機會。

風險傾向與所處環境有關，如果處在相對剝奪的環境裡，也就是自己實際收入遠遠落後在預期收入之下，這時的冒險傾向可能改變，因爲人在損失的環境中的風險傾向高（Levy, 1992），以致於原先也許必須有至少 5%的人參加革命以後，自己才會跟著去革命的人，現在變成只要有 3%參與革命，就足夠使自己也參與革命。風險傾向關乎個人的效用函數，而函數的位置則受到相對剝奪感的強烈與否影響。

與囚徒困境一樣有名的另一個形式理論基礎模式是懦夫遊戲。在這個賽局的情境中，兩人開車相對撞，看誰是怕死先閃開的一方就算輸。和囚徒困境最大的不同在於，當雙方如果都不合作（即不閃開）時，得到的結局對雙方都是最悽慘，而在囚徒困境中，兩人都不合作時，得到的結果對雙方都只是次悽慘的（即中度懲罰），所以兩囚才都有可能選擇不招認。核子戰爭的例子就是懦夫遊戲，亦即一方發動戰爭，另一方也發動，大家相互毀滅（最慘），一方發動，另一方不發動，就是另一方投降或毀滅，但起碼世界不會全部毀滅（不是最慘）。如果把軍備競賽套用囚徒困境理解，是表示雙方理

性的各牟其利，其中有默契與信任的因素；如果套用懦夫遊戲理解，就成爲所謂恐怖平衡，是雙方各避其害，不存在合作問題。

生活當中充滿了小例子，可以用這兩種雛形的形式理論模型解釋。試再舉一例，如學生問老師，期末報告要寫多少頁？老師回答說隨便寫多少頁。同學就會去想，其他同學會寫多少頁呢？如果其他同學寫很長，自己不寫長一點，分數就會低，但因此養成寫長報告的風氣後，以後每個人修每一課都會很辛苦，所以大家又養成默契寫短短的，如果有人將行情破壞，必然遭到譴責。這裡形成的均衡，是多人囚徒困境下的均衡。

囚徒困境和懦夫遊戲是政治科學裡最簡單、最基本的形式邏輯。政治科學所採取的數理模型，常常受制於經濟學家與管理學家已發展出來的數理模型，因爲多數政治科學家沒有很好的能力，去發展在數學上更複雜，但在解釋上卻更具普遍性的模型，於是容易在模型使用時出現削足適履的情形。

五、囚徒困境模型的技術問題

關於這些數理模型會產生哪些問題，可以分成兩方面來談。一是技術方面的問題，一是知識論方面的問題。先談技術方面，包括如何決定一次賽局開始與結束的時間？一次賽局如何算開始？何時算結束？作實際戰爭研究的，因爲戰爭變化快，會以一日爲單位，也可能因應不同變化用別的單位來算，職業籃球的表現以一週或一月爲單位，看一個棒球手一輩子的打擊率變化可能是用一年爲單位。

其次，在囚徒困境中的兩囚是同時出招，而不是先後出招。在實際應用中，同時出招如何界定。許多人擅用囚徒困境分析兩岸關係，其實有很多限制，有的時候，雙方一年還不出手，有的時候可能一天出手兩回，這就要重新改變對出手的定義，或對行爲的分類，比如不出手也算一個行爲，但這就更要靠人爲定義，不能想當然爾。因此形式理論中，什麼行爲算是一次行動或抉擇，不是那麼理所當然就能決定的，研究者必須很武斷地定義出招的行

爲。

再其次，行爲選項要如何歸類？囚徒困境和懦夫遊戲的行爲選擇很簡單，都只有兩項（招或不招；對撞或閃開），但實際的決策沒那麼簡單，一個行動是代表衝突或合作總是說不準，甚至代表衝突的行爲在一段時間後變成代表合作，也不無可能。比如台灣在一九九〇年發表國統綱領，當時北京視爲是一個影射衝突的動作，在二〇〇〇年時，北京則認爲國統綱領的堅持是一個善意的動作。一個動作到底是衝突還是善意（合作），不是囚徒困境或懦夫遊戲假定的那樣好判別。

另外，行爲者在效用函數上呈現的優先序，對判斷他們是在參與什麼性質的賽局很重要。如果雙方衝突的結果對行爲者而言是最嚴重的話，則表示是懦夫遊戲。也就是說，效用函數定義了賽局的種類。依照賽局理論的原本設想，函數是內生的，作爲解釋項，行爲則是依變項，或被解釋項。但實際上，函數的種類往往依靠行爲的表現來類推，此即 revealed preference（Herriges and Kling, 1999），則解釋項的數值來自被解釋項。

同樣嚴重的限制是，行爲的分類往往不清不楚，或依情境而轉變，於是剝奪了研究者藉由行爲反推賽局形式的機緣。甚至，效用函數是一種政策戲碼，刻意表現成某種對結局的偏好優先序，那研究者就更難解釋行爲的發生了。比如，圍堵政策之初，艾森豪用所謂巨棒政策，但戰略家認爲沒有用，因爲假使敵方用傳統武器進行蠶食間歇的攻擊，怎麼可能不分青紅皂白地就用核子武器反擊？這裡值得介紹一個決策模式裡常用的概念叫 last clear chance，亦即最後可以解除危機的能力在誰手中？如電梯要關門了，一人最後跑進來，超載鈴突然響起，誰要出去？當然是他，他造成電梯鈴響，是爲 last clear chance。在實際戰場上，假設我方架設一排鐵絲網，對方一碰到它，核武就自動發射，則萬一造成核戰，對方要負最後責任（Schelling, 1960）。

但對方可以顛覆這個設計，把自己矇上眼衝過來了，這時的責任就難分了，因爲我方可能有即時解除的能力，但也許來不及。到底有沒有及時解除的能力是國家安全的最高機密，表面上一定要假裝沒有這個能力，將後果的

責任推往對方，使對方不能發動衝鋒。形式理論就要去計算，last clear chance 如何設計，如何讓己方的行動到達一個完全不能自我掌控的境界，或近似之，就是所謂的近似值，最後近似到對方沒有辦法在行動上分辨我方還有沒有選擇的餘地時，就達到迴避 last clear chance 的壓力。

艾森豪的巨棒政策就是一套精心設計的表演，他刻意在言談中表現成老粗，假裝不懂蠶食進攻與大規模入侵的差別，使得對方（或美國其他憂心忡忡的戰略家）以爲他真的會訴諸巨棒（Quester, 1979），則 last clear chance 就落到華沙公約組織身上。但艾森豪的表演使得形式理論所賴以建立的效用函數，變得意義模糊。到底艾森豪的表演算是美國的效用函數呢？還是必須另外發掘「真」的效用函數？

決策者所處的環境是，根本不知道對方做一件事是真的或假的，並非事實上不能知道，而是在觀念上，一個官僚決策過程中的真與假之間，區分本不清楚。在知道真假之前，研究者要用什麼方法來判斷，現在是囚徒困境或懦夫遊戲的結構在主導行爲？如果真假之區分本不存在，決策者乃隨時空情境而改變偏好，如此賽局理論的價值等於遭到根本質疑，因此賽局理論的適用，必須建立在對人偏好形式的某種固定假設中。比如人們相信推動台獨中共不會打來，因爲中共會考慮經濟發展，也考慮避免與美國衝突，於是會不了了之。形式理論要求研究者應當計算中共發動攻擊與否的各個可能的不同結局，將之依照效用函數排序。到底最糟的情況是和美國打一仗，還是中國民族主義的崩潰？答案關係到中共玩的是哪一個賽局。如果是囚徒困境，但台灣以爲是懦夫遊戲，於是宣示台獨，中共就會出乎意料衝過來。反之，如果是懦夫遊戲，但僞裝成是囚徒，台灣可能就不會輕舉妄動。至於想測中共處在什麼賽局情境中，不一定測的出來，或測出來也沒有用，因爲情境會轉變，甚或中共自己也不見得事先知道哪一種效用函數會在當下出現，在台灣宣佈獨立後效用函數會有什麼變化──從懦夫變成囚徒，或反之？

這就導出最後一個更嚴重的技術問題，即不但沒辦法客觀上決定效用函數，且當事人的函數是變動的，而非明確一致。解決方法也有，假設中共游

移於囚徒與懦夫兩種效用結構之間，形式理論家發明一種 macro game 或 meta game，乃更高一層的超級賽局，決策者必須先選擇進入哪一種效用函數。決策分成兩層，第一個決定關於當下決策情境是在哪一個效用函數之下，決定了之後，在已經進入的那個模型中發展一套行爲方式。這裡的麻煩是，有幾個效用函數要同時被考慮？有幾個因素在決定中共會進入哪一種效用函數？進入特定效用函數的超級效用函數怎麼決定？

六、囚徒困境的知識論瓶頸

比之技術問題，在知識論方面的難題過之而無不及。首先，囚徒困境中的兩囚，爲什麼要完全站在自己的立場來考慮問題？[6]爲什麼非如此不能稱爲理性？理性的人可否站在團隊的角度考慮問題？囚徒困境的分析當然是假設，人只能站在自己的立場來考慮問題。如果是團體立場，比如國家利益的計算，但這時仍須將國家個體化，形成與其他國家之間區隔，以致於在賽局理論的架構下，不存在單純爲其他國家利益考量的外交理性，更別提爲了其他國家而犧牲自己國家的可能性。數理模式的瓶頸是只針對站在自己立場考慮問題者如何行動提出解釋，但就不能說明不完全站在自己立場考慮問題者如何行動。囚徒困境要成立，首先要考慮的就是這種將研究對象加以個人化的做法，是不是出自於特定的基督教工業文明史，或個人主義現代性的文化假定，因爲他們都是以解放個人於封建、教會、宗族的束縛爲訴求，故爲了集體利益而自我犧牲的行爲者，都是被解放的對象。

其實這個背景是相當明確的。在中國古代的盜墓劇裡，都是父子搭檔，其中的理性因素就在於，父不會殺子，子不會殺父，而倘使不用父子檔時，在墓外等候的一方一時貪婪，很容易可以將搭檔困在其中悶死，自己獨吞墓寶。從囚徒困境來看，盜墓父子不理性，因爲他們的計算單位包括對方的好

6. 參見 Ling, Lily H. M.（2002）書中閒宗明義討論她自己學習囚徒困境之經驗.

處在內。能不能說這對父子有長遠利益，因此彼此合作？可是任何兩個盜墓人都可以發展長期合作關係，何以仍非父子檔不能彼此信任？可見父子關係影響了信任關係，甚至決定了信任關係。換言之，父子盜墓檔的信任關係不是理性抉擇的結果，而是決定了理性計算單位爲何的依據。認爲盜墓父子不殺對方的欠缺理性，是符號邏輯假定的，不是決策者假定的。故謂參與賽局的玩家之間沒有任何情感連繫（關係）的假設，並不正確。

第二個問題是，形式理論家假設每個人都追求得最高分（每個人都是理性的），所以要對於邊際價值高度敏感，因此研究者應該根據效用函數去計算每一點的斜率，找到與環境制約的切點。[7]但就連Riker本人都會承認，這假設有問題。Riker發現，選舉時總是還會有人去投第三黨的候選人，即使明知必落選卻仍堅持。他因此主張，對這些怪人而言，投票的過程比投票的結果更重要（Riker and Ordeshook, 1968），不過他認爲這種情形很少見。Riker可能低估了過程的重要性。

是少數人在少數場合如此，還是絕大多數人都認爲過程比結果更重要？投票的動作，使人和社會上其他人取得政治認同上的聯繫，此地出現的正是形式理論家迴避的「集體」與「個體」之間的認同／角色／意識問題。換一種問法，即有目標的感覺比較重要，還是達到目標比較重要？是取得勝利較重要──個體化思考？還是透過行動使自己在社會過程中有參與的角色更重要──角色化思考？

形式理論碰到另一個知識論上的問題，是對於利己與利他之間的差異無法處理。[8]故會認爲在公車上讓座的人，是在利己，因爲讓的感覺讓自己很舒服。這時讓座者的目標是讓別人達到舒適的目標，即以別人的目標爲目標，所以仍屬於目標導向。這種不能利他的規定，也反映了前面提的關於社會關係的瓶頸，故母親對孩子或軍人對國家的犧牲，都只能是利己的行爲。然而，

7. 這一點評論適用於連續函數，不適用於不連續函數。

8. 強有力的批判見 Gilligan, Carol (1982).

公車上碰到的是不特定的人，什麼人的目標會成爲是自己利他時要達到的目標，沒有定論，除非形式理論接受社會關係制約了利他行爲，既然利他被當成利己，則社會關係也就制約了利己的行爲，使社會關係在利己的計算在開始之前，就已經決定行爲者的偏好傾向。

第三個問題是，形式理論所預期的計算，是天生的，還是學習（有人教過）後才會的？形式邏輯能否判斷，一些看似目標導向的行爲，是因爲行爲者學習了目標導向的文化，亦即是後天教出來的，還是不教也會。假如是教出來的問題就大了，這裡的意義有兩層：一是學會了目標導向的理性，凡事都照這個原則行爲；二是行爲者另有行爲動機，但由於只學過目標導向的理性，故只能用目標導向的語言表達自己的行爲。不論是哪一者，形式理論的解釋都變成知識論層次上的套套邏輯。

最後一個問題，即語言是不是代表現實（Gefwert, 2000）？形式邏輯講的是符號，簡潔地翻譯日常語言來分析行爲。但語言似乎並不是用來指涉現實的工具，而是在被用的過程中與情境結合。社會的實際是什麼，語言在其中扮演了關鍵的角色。簡言之，語言不是用來指涉實際的物或動作，而是參與了創造並組織行爲者對於實際的感覺，所以語言學家與現象學家都稱語言爲一種遊戲。沒有語言，就沒有當下的情境，也沒有現實，蓋語言是社會實踐的結果。如此一來，效用是語言構成的，並不是外於語言的客觀存在。

形式邏輯認爲所有的語言都能化成邏輯符號，所以可以用形式邏輯來解釋人類行爲。這種知識論的前提，是以爲日常所使用的語言，是用來指涉與語言無關的外在世界的工具，只有這樣的假設，才會認爲可以把語言進一步轉化爲邏輯，而且用它來解釋社會實踐中的行爲。

七、作爲規範的形式理論

形式邏輯雖是演繹法，且和統計分析屬於歸納法不同，但兩者的更大共通性，在於都認定符號或所定義的變數本身，是可以和社會實際無關，或在社會實際之外來指涉客觀、特定研究對象的工具。形式邏輯發展到今天，成爲最先進的一種政治科學模型，在學習使用此模式的過程中，不能忽視研究者是被教會的，研究對象也是被教會的，形式理論是政治科學界與政策分析界的規範，故所起的教導作用比分析作用更大，自也有文化改造的作用，這個在知識論上的自我證成，是已經進入形式理論思惟模式的研究者所反省不了的。

參考書目

Barry, B., and R. Hardin. 1982. *Rational Man and Irrational Society*. London: Sage.

Brodbeck, May. 1968. *Readings in the Philosophy of Social Sciences*. London: Macmillan.

Cline, Ray S. 1977. *World Power Assessment 1977: A Calculus of Strategic Drift*. Boulder, Colo.: Westview Press.

Gefwert, Christoffer. 2000. *Wittgenstein on Thought, Language and Philosophy: From Theory to Therapy*. Aldershot, UK: Ashgate.

Gilligan, Carol. 1982. *In a Different Voice*. Cambridge: Harvard University Press.

Gurr, Ted. 1970. *Why Men Rebel*. Princeton: Princeton University Press.

Herriges, Joseph A., and Catherine L. Kling. 1999. *Valuing Recreation and the Environment*. Cheltenham: E. Elgar.

Huang, Jing. 2000. *Factionalism in Chinese Communist Politics*. Cambridge: Cambridge University Press.

Kaplan, Morton. 1957. *System and Process in International Politics*. New York: John Wiley.

Levy, Jack S. 1992. "Prospect Theory and International Relations." *Political Psychology* 13, 2: 283-310.

Ling, Lily H. M. 2002. *Postcolonial Learning between Asia and the West: Conquest of Desire*. New York: Palgrave.

Little, I. M. D., and M. F. G. Scott. 1976. *Using Shadow Prices*. London: Heinemann.

Olson, Mancur. 1965. *The Logic of Collective Action*. Cambridge: Harvard University Press.

Parsons, Talcott. 1969. "On the Concept of Political Power." in Bell, Roderick, Edward, David V., and Wagner, Robert Harrion eds. *Politics and Social Structure.* New York: Free Press.

Quester, George H. 1979. "Was Eisenhower a Genius?" *International Security* 4, 2.

Riker, William. 1995. "The Political Psychology of Rational Choice Theory." *Political Psychology* 16, 1.

Riker, William, and Peter C. Ordeshook. 1968. "A Theory of the Calculus of Voting." *American Political Science Review* 62: 25-42.

Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge: Harvard University Press.

Stokey, Edith, and Richard Zeckhauser. 1978. *A Primer for Policy Analysis*. New York: W. W. Norton.

Howey, Richard S.著。1999。《邊際效用學派的興起》。晏智杰譯。北京：中國社會科學出版社。

徐振國。2002。〈政治學方法論偏頗發展的檢討〉。《政治與社會哲學評論》2：123-179。

The Epistemological Limit of Formal Theory in Political Science —Game Theory Revisited

Chih-yu Shih*

This paper discusses the epistemology of modeling in political science. It touches upon the notions of utility, marginal value, shadow price and their application as well as limitation. The paper takes particular interests in prisoners' dilemma and its limitation.

Key words: formal theory, game theory, prisoners' dilemma, shadow price, marginal utility, political science methodology

* Professor, Department of Political Science, National Taiwan University, Taipei.