

行政院國家科學委員會專題研究計畫 成果報告

為何邏輯推論那麼難?假設檢定歷程中的歸納與演繹成分的分析探討

計畫類別：個別型計畫

計畫編號：NSC92-2511-S-009-015-

執行期間：92年08月01日至93年07月31日

執行單位：國立交通大學工業工程與管理學系

計畫主持人：洪瑞雲

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 8 月 9 日

行政院國家科學委員會專題研究計畫成果報告

合作與解釋影響科學推理作業的機制：語文資料分析

計畫編號：NSC 92-2511-S-009-015

執行期限：92 年 8 月 1 日至 93 年 7 月 31 日

主持人：洪瑞雲¹

計畫參與人員：黃文毅¹ 黃富源¹

¹國立交通大學工業工程與管理學

為什麼邏輯推理這麼難：假設檢定過程中歸納與演繹成分的分析探討

一、摘要

科學發現的工作是在對觀察到的現象導出其規則性。但此一「若 p 則 q」的法則只能以證偽邏輯來驗證，在沒有「若 p 則~q」的反例時，此法則才能暫時被保留。然而，一個由經驗導出的法則（假設）是充滿了意義的；而邏輯證明只是抽象的形式演算；人真能只以「形式邏輯」來驗證一個「經驗」法則嗎？我們推論人不易使用證偽邏輯是因為人無法完全區隔法則中之歸納與演繹兩種成分。假設檢定過程中演繹的形式邏輯不會是人唯一的考量，此形式邏輯所規範的變項間的經驗意涵也是必要的考量。由 24 為大學生與 20 為有研究經驗的研發人員在法則檢定作業上的表現顯示，法則中的 p 與 q 間經驗上的關連意義是影響法則檢定正確率的重要因素。

關鍵字：邏輯推理、假設檢定、歸納、演繹、科學發現、解釋

Abstract

To explain why people often test a hypothesis/rule by confirmation strategy, rather than falsification strategy, intrusion of inductive reasoning component in deductive reasoning task was examined. Twenty-four college students and first year master program's students and 20 R&D workers participated in the study and solved 20 four-card selection tasks differing in the concreteness of the rule description. Results show that when rules were embedded in concrete but meaningless stories, R&D workers' performance suffered more than college students. Meaning which is induced from experience thus seems plays a significant role in deductive reasoning tasks.

Key words: inductive reasoning, deductive reasoning, hypothesis testing, scientific explanation.

二、緣起與目的

科學家的主要工作在為我們觀察到的經驗現象提出一個描述性的解釋，或為現象背後的共通法則找出一個運作機制的解釋 (Simon, 2000)。科學發現有賴兩套思考推理的方式：演繹推理與歸納推理 (Dewey, 1933/1991)。演繹推理是根據一些普遍的原則去推知某些特定的例子的值或屬性的一種方式。歸納推理則是一種根據“經驗”去判斷某些現象背後共同的法則的一種推論方式。F. Bacon (1561-1626) 因此稱歸納思考為科學的基本方法。

科學解釋的先決條件在發覺不同事物間是否有共同的地方，因此，科學家首要的任務是要仔細觀察事物所呈現的各種面向，找出並描述它們的共同面向，進而解釋造成此現象的因果機制。然而，經歸納推理而得到的法則/結論只能稱之為假設，因為此結論乃是根據有限的案例建立起來的，其普遍性、正確性必須以演繹邏輯的方式加以測試。在邏輯上，經歸納而得的法則，即使得到再多新案例的支持，亦無法證明其真實性。相反的，萬一有任何與此法則衝突的案例出現時，則此法則的正確性便可輕易的被推翻了。Popper (1959) 因此主張由歸納而得的假設必須以「證偽」(falsification) 邏輯的方式測試。在無法被證偽的情形下，此假設才可能成為一個可靠的知識。科學家的法則發現過程中因此需要同時運用歸納與演繹兩種推理方式。

Simon & Lea (1974), Klahr & Dunbar (1988) 及 Klahr & Simon (1999) 曾分別以兩空間理論來看說明科學法則發現的歷程。根據他們的理論，法則發現包括：1. 在面對有限的案例去推測其背後的法則時，受試者的工作相當於在一可能的法則或型態的集合中去尋找、選擇一個適用於解釋已知案例的法則；2. 假設驗證的過程中則相當於是設法尋找或產生新的案例以增加線索以利型態的辨識、或法則的判斷。假設形成是一種概念形成、或型態辨識的工作 (Bruner, Goodnow, & Austin, 1986)。當事人是根據其已有的相關知識架構對觀察到的現象間的共同性提出一個分類或解釋的抽象架構 (概念型態的辨識)，以涵蓋、說明這些現象。現象的存在可說是客觀的、獨立於個人之外的。但一個人是否可以觀察到現象及現象間的共同性，甚至提出一套自己的解釋架構則是視那個人是否具有相關的知識背景並能由有限的案例中萃取出可作為解釋架構的有意義型態而定。經驗與意義度因此是法則發現的重要依據。

相對於 Popper 所稱的證偽策略，研究指出，人在假設檢定時有極強烈的證真偏好 (Wason, 1960, 1968; Wason & Johnson-Laird, 1972; Evans, 1989; Newstead & Evans, 1995)。大多數人從事假設檢定工作時，傾向於以符合自己假設的正例來進行測試的工作。在這種證真策略下，我們雖可得到許多肯定我們假設為真的證據，但卻無法發現假設當中可能有的缺陷。以 Wason (1960) 所設計的翻卡片 (four-card selection problem) 作業為例，在測試「若 $p \rightarrow q$ 」的法則時，

p , q , $\sim p$, $\sim q$ 四張卡片中, 測試 p 的結果即使得到 q , 也無法證明此法則是有被遵守; 但若測試 p 而得到 $\sim q$ 或測試 $\sim q$ 的結果得到 p , 此法則的正確性即可被推翻掉。因此, 在翻卡片作業中正確的反應應是選擇測試 p , 及 $\sim q$ 兩張卡片。但是一般大學生選擇同時測試 p , 及 $\sim q$ 兩張卡片的情形卻遠低於 20% (Wason & Johnson-Laird, 1972; Evans, 1989) Oaksford & Chater (1994) 的後攝分析則指出典型的卡片選擇是 p (89%) > q (62%) > $\sim q$ (25%) > $\sim p$ (16%), $\sim q$ 的選擇嚴重偏低, 顯示一般人很不容易由證偽的觀點去檢定一個假設。

人偏好證真的假設檢定策略的可能理由之一是假設檢定作業中含有歸納推理的成分, 不完全是演繹邏輯的作業, 本研究的主要目的即在探討假設檢定作業中歸納與演繹推理相互影響的方式。

假設產生與假設檢定並非是完全獨立的兩個歷程。一個法則的發現可能需要許多提出假設、測試假設、根據測試結果修訂假設或提出新假設的迴路。因為在科學的研究中我們永遠無法知道真正的法則為何, 因此, 不斷的證偽的工作在假設檢定上是十分重要的。假設產生與測試既是科學思考的核心, 是人類發現新的知識法則的主要方式, 許多研究便在探討如何幫助人去除證真偏好、有效的使用證偽方式去測試其假設。研究發現, 直接教導證偽邏輯的效果並不好 (如, Gorman, Stafford, & Gorman, 1987)。然而, 只要題目是受試者熟悉的題材, 則證偽似乎就相當容易, 顯示經驗是很關鍵的因素 (如, Kareev &

Halberstabt, 1993), 而經驗通常是來自歸納的結果。

Oaksford & Chater(1994)的資訊增加理論 (information gain theory) 認為選卡片作業中的四張卡片可被視為是龐大母體中的四個抽樣樣本, 一般人演繹推理的表現不理想, 可能是因為他們將選卡片作業當成是抽樣的問題, 而使用歸納推理的結果。Cheng (1997)也認為以歸納推理的方式推論某規則 (如 $p \rightarrow q$) 為真的可能性, 各種情況在母體中佔有的比例 (如, p 、 $\sim p$ 、 q 、 $\sim q$ 各有多少張, 各佔多少百分比) 會影響受試者評估該規則為真的機率。以規則「只要是女人 (p) 就有葡萄酒喝 (q)」為例, 如果全部有 10 人, 其中 2 個是女人 ($p = .2$), 8 個男人 ($\sim p = .8$), 此時, 若(1) 10 杯飲料中有 9 杯葡萄酒 ($q = .9$), 1 杯白開水 ($\sim q = .1$), 或(2) 10 杯飲料中有 3 杯葡萄酒 ($q = .3$), 7 杯白開水 ($\sim q = .7$); 則要驗證「女人就有葡萄酒喝」的假設時, (1)中女人(p)與葡萄酒(q)的比例為 2:9, (2)中女人(p)與葡萄酒(q)的比例為 2:3, 「女人就有葡萄酒喝」在(1)中規則為真的機率顯然大一些。由此推論, 如果人在做選卡片作業中有用到歸納推理的成分, 那麼 p 、 $\sim p$ 、 q 、 $\sim q$ 各有多少張, 各佔多少比例, 將會影響著受試者選擇測試的卡片。

Kirby (1994) 以信號偵測理論 (signal detection theory) 的角度分析選卡片作業。他認為如果受試者存心尋找反例, 他會將反例(測試 p 而發現其反面為 $\sim q$, 或測試 $\sim q$ 而發現其反面為 p 視為一個「命中」(hit); 相反的, 若雖然 p 的反面是 $\sim q$, 但受試者卻沒有

選擇測試 p ，則稱為「錯失」(miss)；未測試 p 卡片，而且 p 卡片背後也只是 q 時，則是個正確的拒絕 (correct rejection)；選擇測試 p 卡片，結果反面是 q ，則是個假警報 (false alarm)。Kirby 認為命中、錯失、正確拒絕、假警報的成本和利益，會影響測試一張卡片時的效用，進而影響人的選擇。他認為選卡片作業中人偏好選擇 p ，是因為選擇 p 卡片的效用，大於不選擇 p 的效用。他分析人選擇測試 p 的情況有二：

1. 就機率而言，當 $\sim q$ 的機率增加時，人測試 p 命中 $\sim q$ 的機率較大，所以測試 p ；
2. 就效用而言，當測試 p 命中 $\sim q$ 的利益提高、或正確拒絕測試 p 的獎賞減少、或沒有測試 p 而錯失 $\sim q$ 的成本增加、或測試 p 發現是假警報 q 的成本減少時，人會較傾向選擇 p 。

Kirby 理論的前提假設是人會使用反例 (p 和 $\sim q$ 的配對) 來測試法則，且測試 p 和 q 的成本效益會影響測試的行為。相反的，Oaksford 和 Chater (1994) 則認為在日常生活中一般人驗證假設時並不追求邏輯上的正確性，而只要相對上可以提高該假設正確的機率就可以了。此外，人生活中的推論是用生活中大量的知識作為推論的基礎。人的生活的知識是如此龐大，記憶中相關的資訊往往可以幫助思考的進行。他們因此認為假設檢定過程中人會藉由經驗去評估選擇測試 p 、 $\sim p$ 、 q 、 $\sim q$ 的結果對 $p \rightarrow q$ 的真偽所能提供的資訊量的多少，如 $\sim q$ 在母體中的機率小時，測試結果所提供的資訊會較大，人就會選擇測試 $\sim q$ 。

林奕祺 (2002) 為了檢定究竟訊

號偵測理論或資訊增加理論何者相對上比較符合實際的選卡片行為，他以 34 個台灣的大學生為對象，操弄指導語中是否有提示法則被違反，以及 p 及 q 的機率大小，結果發現只要指導語中有提示法則可能被違反，不論 p 及 q 的機率大小，受試者同時測試 p 及 $\sim q$ 的情形就上升(有提示法則可能被違反, .25；沒有提示, .16)，顯示大學生是有使用證偽邏輯的能力的，但只有在他們懷疑法則可能被違反時才會使用。就 p 及 q 的機率大小所產生的影響而言， p 的機率小時，受試者測試 p 的次數顯著上升； q 的機率大 ($\sim q$ 的機率相對上較小) 時，測試 $\sim q$ 的次數也顯著上升；此結果除了為資訊增加理論提供額外的支持外，也顯示假設檢定過程中除了演繹推理外，還含有歸納推理的成分。

最近，Morris (2002) 以 30 個大學生為對象的研究指出，雖然一般的數學證明題解題時需要用到的是演繹邏輯的推理方法，受試者在證明如「證明對每一正整數 N 而言， $N^2 + N$ 為偶數」這類問題時，有 40% 的受試者會認為只有用歸納推理的方式 (尋找正例) 才可證明此命題的真偽，有 30% 的受試者認為歸納與演繹推理同樣可證明此命題的真偽，但又認為演繹推理的證明方式優於歸納推理；只有 30% 的受試者認為只有演繹推理可證明此命題的真偽，而歸納推理不能。然而，受試者對歸納與演繹在數學證明上的效力所持的這些信念與他們其實際上的解題表現並無相關。Morris 因此推論影響數學推理的能力有兩種，一為個人是否有掌握到問題中抽象的「前提 - 結論」間關係，並據此

從事形式思考、演算的能力；另一為個人對數學中的物件（如 1、2、3、4 等數字及加、減、乘、除等演算）的本質及其規則性的信念，如，數字代表什麼？何為數字系統等。我們因此推論，數理邏輯所規範的法則如「 $p \rightarrow q$ 」其意義也有兩層，一為純粹的形式關係及此形式關係所允許的演算、演繹推導；另一為此「 $p \rightarrow q$ 」是否真能由的經驗加以佐證的信念。人在假設驗證時也就很可能同時會受歸納思考（「 $p \rightarrow q$ 」是否有經驗上的意義）的影響而出現證真偏好的現象。

本研究的目的是探討人在假設檢定歷程中受演繹推理與歸納推理影響的方式及程度。我們推論，由於一個假設或法則的本質是在對真實世界加以描述或解釋，其真實性或意義度基本上是植基於證據的歸納，尤其是正面可以支持此假設可以成立的證據上。因此，在假設檢定過程中歸納推理與演繹推理必須交互使用，完全的形式推演對假設檢定而言是不夠的，考慮假設的內容的經驗意義因此是受試者必然的反應。我們將以選卡片的作業為假設檢定作業，我們預測在內容具有經驗意義的題目上（如附錄一），將可觀察到受試者以歸納的方式進行假設檢定的行為。在內容不具有經驗意義的題目上，由於假設的內容沒有經驗上的實質意義，歸納推理的干擾將下降，但由於假設的內容缺乏經驗的佐證，證偽邏輯演繹的正確性並不會因此上升。由於經驗是和年齡與工作內容相關的，因此我們也預測，假設檢定中歸納推理的成分並不會因年齡、或相關假設檢定的工作經驗增加而降低。我們的受試者將分成

無研究經驗的大學生及有研究經驗的研發人員。我們預期此兩群經驗不同的受試者在假設檢定過程中都同樣會表現出歸納與演繹兩種推理成分。

三、方法

受試者及年齡/經驗的分類

本研究的受試者分別為 24 個大學生及碩士新生、及 20 個有碩士以上學位且從事研發的人員，其中 16 位研發人員係來自工業技術研究院，一為大學教授、三位為大學博士班研究生，主要皆為理工背景。這兩群有、無研發經驗的受試者年齡，平均差距約 10 歲（表 1）， $t(20.34) = 6.39, p < .001$ 。

表 1: 年齡

	研發 經驗		
	無	有	全體
<i>M</i>	21.46	33.95	27.14
<i>SD</i>	1.77	8.59	8.60
最小值	19	23	19
最大值	22	52.00	52

假設檢定作業及經驗上意義度的設計

由於在林奕祺（2002）的研究中並沒有觀察到受試者會使用問題中的數字及機率的演算，因此，我們在問題中並沒有加入機率的操弄，而只操弄問題在經驗上的意義度，以觀察受試者在假設檢定時的考量因素是否會因此而不同。假設檢定作業以 Wason 選卡片作業為設計依據，每一題均含一「若 p 則 q 」的法則，受試者的工作在驗證此法則是否為真。

有實質經驗上的意義的題目。

1 題練習題及 16 題測試題屬此類「若 p 則 q」的法則中的 p、q 兩個變項皆為有實際經驗內容的物件 (object)，且 p 與 q 的條件式具有實質上的意義 (有經驗上的基礎)，但「若 p 則 q」的法則本身則是虛構的。如，「若開車 (p) 則年齡皆是 25 歲以上 (q)」

無經驗上的意義的抽象題。

「若 p 則 q」的法則中的 p、q 兩個變項雖皆為經驗內容的物件，如，英文字母、松樹等，但在實際經驗中 p、q 並無任何關係。在問題中，p、q 二者的關係是人為的設定，沒有經驗上必然的共變 (covariation) 或合理性。如，「若卡片的一面藍色 (p) 則反面是個圓形 (q)」。

抽象法則。「若 p 則 q」的法則中 p、q 項為抽象符號，如「若甲則乙」。

假設檢定的相關應變項

演繹邏輯。「若 p 則 q」的法則意涵「若 $\sim q$ 則 $\sim p$ 」，測試「若 p 則 q」的法則時，只有當測試 p 而得到 $\sim q$ ，或測試 $\sim q$ 而得到 p 的時候才是此法則為偽的證據。因此選卡片作業中給定的四個選項 (分別對應到 p, $\sim p$, q, $\sim q$) 正確的選擇是同時測試 p 及 $\sim q$ 兩張卡片。

歸納推理。「若 p 則 q」的條件句所表達的是 p 與 q 間的充分關係，此充分關係可否成立完全視真實世界中是否有支持此共變關係的證據。因此，當代入任何代表真實現象的名詞 (概念) 時，「若 p 則 q」的條件句所表達的充分關係不一定皆可為

經驗所支持。如，「如果花是紅色的，則花是無毒的」、「如果汽車發不動，則是汽車生氣了」。在閱讀完一個具象的條件句之後，我們將要受試者以 9 點量表評量此條件句的合理程度以衡量受試者認為此條件句是否可被其經驗所佐證。此外在選擇測試方案後，並由受試者對其選擇的測試方法的信心程度以 0 致 100 的百分比加以評估，以檢視其中是否有歸納的成分。因為演繹邏輯的結果是真或偽應具有 0 或 1 的全有或全無的信心，偏離此二點的信心評估值顯示受試者並未完全使用演繹推理。我們也將蒐集受試者解題時的口語資料，由口語分析中找尋受試者對此法則的評論及測試時的行為，以進一步了解歸納推理如何影響此法則的測試。

實驗設計

研究經驗與題目在經驗上的意義兩個自變項形成一 2×2 的實驗設計。經驗為組間變項，題目在經驗上的意義為組內變項。

過程

以個別實驗的方式進行。大學生是在學校的實驗室中進行，研發人員則在其工作地點的研究室進行。受試者在 1 個練習題之後，先接受 16 個在經驗上有意義的假設檢定作業，接著是 3 題經驗上沒太多意義的抽象題，最後則是 1 題「若甲則乙」的法則題。在每一題中，受試者分別要評量此問題中的法則的合理程度 (1 至 9 分)，然後由給定的 4 個案例 (對應至 p, $\sim p$, q, $\sim q$) 中選擇適當的測試法則真偽的

方式，可複選，且要對每一選擇的測試結果是否可以提供法則真偽的依據下一個 0 到 100 分間的信心評估。受試者同時被要求在解題過程中將其思考的內容說出來。實驗全程錄音。整個過程約需時 1.5 ~ 2 時

四、結果與討論

測試 p 及~q 的平均次數及比率

由於問題是否具有具體的經驗意義（有、無）與研發的工作經驗（有、無）二因子重複量數變異數分析結果顯示二者的交互作用對正確選擇 p 與 ~q 無影響，因此我們直接以問題是否具有具體的經驗意義或研發的工作經驗為自變項進行單因子的平均數差異檢定。全體受試者在 17 題具體題（含練習題）3 題題目無經驗意義的抽象題及法則題上正確的測試如表 2 所示。

表 2：測試 p 及~q 的平均次數及比率

	具體		抽象		總分
	具體	抽象	法則	總分	
<i>M</i>	4.07	0.50	0.32	0.82	4.89
%	.24	.17	.32	.20	
<i>SD</i>	3.14	0.73	0.47	1.11	3.54

以 t 檢定比較有無研究經驗的兩群人在這些測試題上的表現的差異（表 3），結果如 3 所示，兩組不同研究經驗的人在具體題上的表現相似，在直接用法則表示法呈現的抽象題中二組表現也無差異，但若問題是以無經驗意義的名詞間的關係呈現時，年輕無研究工作經驗的學生的正確率則顯著高於年紀大他們約 10 歲的研發人

員，顯示有工作經驗的研發人員在面對無經驗基礎的問題時，提取及運用邏輯推理的程度會受到限制。

表 3：兩組同時測試 p 及~q 的 t 檢定

	具體	抽象	法則	抽象	總分
	17 題	3 題	1 題	總分	
沒研究經驗					
<i>M</i>	4.38	0.71	0.42	1.13	5.50
<i>SD</i>	2.87	0.81	0.50	1.19	3.43
%	.25	.24	.42	.28	.26
有研究經驗					
<i>M</i>	3.70	0.25	0.20	0.45	4.15
<i>SD</i>	3.25	0.55	0.41	0.89	3.62
%	.21	.08	.20	.11	.20
<i>t</i>	0.71	2.16	1.54	2.10	1.27
<i>df</i>	42	42	42	42	42
<i>p</i>		.04		.04	

若直接以單因子重複量數的方式檢定具體題與抽象題正確選擇 p 及~q 的平均比率時發現，17 題具體題與 3 題有故事內容但無經驗基礎的抽象題間的答對比率差異接近顯著程度， $F(1, 43) = 3.14, .05 < p < .08$ 。17 題具體題與抽象法則題間的答對比率則差異不顯著， $F(1, 43) = 1.20$ 。3 題有故事內容但無經驗基礎的抽象題與法則題間的答對比率差異則顯著， $F(1, 43) = 7.99, p < .01$ ，顯示一般人在法則的表達法上可以正確的運用邏輯推理的程度較高（ $M = .32; SD = .47$ ），當問題是以沒有經驗基礎的故事呈現實，使用邏輯推理的程度會顯著受限（ $M = .17; SD = .24$ ）。由此推論，經驗基礎是推理是否可以符合邏輯的重要條件之一。

相關係數

年齡與 17 題具體題的正確率間的相關不顯著 ($r = -.13$)，顯示只要在問題中的法則是以有意義的故事陳述時，人的邏輯推理正確度不會受年齡影響。相反的，年齡與與 3 題抽象題的相關為 $-.36$ ($p < .02$)，與 1 題法則題的正確率相關為 $-.38$ ($p < .01$) 與抽象題總分的正確率相關為 $-.40$ ($p < .01$)，顯示在抽象的問題陳述中邏輯推理表現隨年齡遞減，極可能是因為隨著年齡上升，經驗與知識也上升，人在推理時依賴經驗的程度也上升了。

口語資料分析

尚在進行中。但由實驗過程中的觀察已發現，法則測試中受試者通常會考慮 p 、 q 、 $\sim p$ 、 $\sim q$ 在母體中的分配機率，再由抽樣成本考量要抽查那一筆資料，如「開車 22 歲以上」，受試者會選擇測試 16 歲者 ($\sim q$) 或不開車者 ($\sim p$) 時的解釋之一是相對上 22 歲以上者 (q) 或開車者人 (p) 數較多，顯示 p 與 q 的經驗上的內容是影響法則檢定的因素。

五、結論與自評

由經驗會影響法則檢定的邏輯作業表現，本研究發現，人的邏輯思考出錯的理由之一是來自 p q 法則中 p 與 q 的內容及 p q 關係是否具有經驗上的意義，此現象顯示人的思考主要需有經驗、意義為基礎。純粹形式邏輯的推演在抽象符號的層次上或許是可行的，但若以具體的名詞取代抽象符號時，即使是有相當的研究經驗的

人，亦可能因名詞的語意內容的干擾而降低了邏輯推理的正確性。我們希望完成口語的資料分析後，能對演繹與歸納思考相輔相成或相互干擾的現象有更進一步的發現。

六、參考資料

林奕祺，2002，假設檢定歷程中機率模式的探討 交通大學工業工程與管理學系碩士論文

Bruner, J. R., Goodnow, J. J., & Austin, G. A. (1986). *A study of thinking*. New Brunswick, NJ: Transaction, Inc. Cacioppo, (1996).

Cheng, P. W. (1997). From covariation to causality: A causal power theory. *Psychological Review*, 104, 367-405.

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, 99, 365-382.

Dewey, J. (1933/1991). *How we think*. Amherst: Prometheus.

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica*, 43, 11-21.

Evans, J. St. B. T. (1989). *Biases in human reasoning: Causes and consequences*. Hillsdale, NJ: Erlbaum.

Evans, J. St. B. T. & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology*, 64, 391-397.

Gorman, M. E., Stafford, A., & Gorman, M. E. (1987). Disconfirmation and dual hypotheses on a more difficult

version of Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology*, 39, 1-28.

Kareev, Y., & Halberstap, N., & Shafir, D. (1993). Improving performance and increasing the use of non-positive testing in a rule-discovery task. *Quarterly Journal of Experimental Psychology*, 46, 729-742.

Kirby, K. N. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection Task. *Cognition*, 51, 1-28.

Klahr, D., & Dunbar, K. (1988). Dual space in scientific reasoning. *Cognitive Science*, 12, 1-48.

Klahr, D., & Simon, H. A. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin*, 125, 524-543.

Morris, A. K. (2002). Mathematical reasoning: Adults' ability to make the inductive-deductive distinction. *Cognition and Instruction*, 20, 79-118.

Oaksford, M. & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608-631.

Oaksford, M. & Chater, N. (1995). Theories of reasoning and the computational explanation of everyday inference. *Thinking & Reasoning*, 1, 121-152.

Oaksford, M. & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review*, 103, 381-391.

Oaksford, M., Chater, N., & Grainger, B. (1999). Probability effects in data selection. *Thinking & Reasoning*, 5, 193-244.

Popper, K. (1959) The logic of scientific discovery. New York: Basic Books.

Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed/) Knowledge and cognition. Hillsdale, N.J.: Lawrence Erlbaum, 105-127.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12, 129-140.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20, 273-281.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. London: Batsford.