教育實踐與研究
2007 年 3 月，第 20 卷第 1 期，119-146

從道德演化的「賽局模型」與「名譽模型」談道德教育－Frank 道德情緒理論之應用

張榮富*

本文引介經濟學者法蘭克（Frank, 1988）道德情緒（Moral Sentiments）理論的兩個模型：以演化賽局為基礎的「承諾模型」與以心理學的配合法則（Matching Law）為要素的「名譽模型」，並且運用這兩個模型來討論德行（Virtue)層面的道德教育，及其在社會發展上所扮演的角色與價值。首先，道德教育在承諾模型的演化賽局中，被定義為：借由增強內在德性，人為的而且短期的改變合作者（好人）與背叛者（壞人）的人數比例，使合作者的人數增加。本文由模型中推導得出：一個社會由低度道德進步至高度道德的過程中，好人與壞人的平均獲利和社會整體的總利益皆會增加，不過此時好人的平均獲利會由大於壞人反轉成小於壞人。道德教育能帶來社會利益卻不能帶來社會正義。其次，本文接續名譽模型的論點推導得出，在道德教育能增強「名譽傳訊現象」時，即當名譽成了內在道德情緒的可靠外在「傳訊」特徵時，好人的比例在長期中會「自然」增加，而且社會整體的總利益也將增加。最後，本文提出「道德教育的內容不限於只在教道德」的兩點技術層面的意見供大家思考。

關鍵詞：道德情緒、道德教育、演化賽局、德行、名譽

* 張榮富：國立臺北教育大學社會科教育學系助理教授
jfchang@ms4.kntech.com.tw

Research Article
Journal of Educational Practice and Research
March 2007, Vol. 20, No. 1, pp. 119-146

A Discussion of Moral Education from Evolutional Game Model and Reputation Model: An Application of Frank's Moral Sentiments Theory

Jung-Fu Chang*

This study introduces Robert H. Frank's moral sentiments theory (1988), and uses it to discuss the role of moral education. Firstly, following Frank's commitment model and its evolutionary game process, moral education is defined as temporarily strengthening human inner moral emotion to increase the number of cooperator (good behavior) in a society. Two conclusions are found. (1) In a low moral society, the defector's benefit is smaller than cooperator's. In contrast, in a high moral society, the defector's benefit is larger than cooperator's. (2) Moral education (improving a society's moral level) does not bring social justice. However, it does bring social benefit. Secondly, following Frank's reputation model, we argue that because reputation can be a strong signal of individual's inner virtue, moral education can strengthen individual's reputation and increase the number of cooperator in a society. In the end, the study suggests that moral education does not need limitation in teaching morality.

Keywords: ***virtue, moral emotion, moral sentiments, moral education, evolutionary game***

* Jung-Fu Chang: Assistant Professor, Department of Social Education, National Taipei University of Education

從道德演化的「賽局模型」與「名譽模型」談道德教育－Frank 道德情緒理論之應用

張榮富

緣　起

在與第一線小學老師的座談中，我們常發現施教者對道德教育，常產生下列兩種困惑。第一、施教者如果注重教導規範性的道德行爲，強調學生應該或不應該做什麼，在教育過程中，施教者的引導很容易流於形式上的規範的要求，而在一個規範價值易隨社會變遷而改變的當代社會中，規範性的道德認知易因受教者日後環境改變，而與新環境中的規範性道德起衝突。受教者日後有可能產生不知所措的現象或產生道德的虛無感。第二，老師注重學童的規範性的道德行爲時，學童易感受到來自老師物質或心理上的報酬或懲罰。遵守道德規範與否，易流於只是追求眼前的報酬或逃避立即的懲罰，這可能只是反映學童的成本效益的考量，有可能是更加強功利取向的價值觀。施教者可能會懷疑眼前品行良好的學生，是否只是更追求眼前報酬或逃避立即懲罰的聰明投機者。是否我們可能反而教育出「品格」與「行爲」不一致的學童？雖然很多老師希望能教導出具有內在德性（Virtue）的人，然而德性層面的道德教育要如何落實似乎是許多小學老師共同面臨的問題。

壹、前言

道德可分爲兩個層面，規範倫理（Morality）層面與德性（Virtue）層面。規範倫理層面的道德是外在制度與倫理形成對人的行爲之他律性規範，會隨着外在社會變遷而面臨挑戰與改變。而德性層面的道德則屬於內在的道德情操之自律性行爲（程亮，2004）。由這兩個層面衍生而來，討論道德教育亦可分爲有關外在的規範性道德教育

與有關內在的德性（Virtue）[1]道德教育兩個大方向。規範倫理層面的道德教育其重點與目的在於教導一個人如何去認知或判斷明確的社會道德規範，並能以有效的行爲去實踐這些規範。德性層面的道德教育則側重在「增强內在德性」，教育人們做具有美德（Virtue）的人（王海明，2001）。

源自於亞理斯多德（Aristotle）思想的德行倫理學（Virtue Ethics）對道德教育的理念，隱含著一個假設，即一個人內在的德性（Virtue）乃是他的行爲長期遵守或違背道德規範所得到的結果。對此，古希臘大師亞理斯多德（Aristotle）清楚的指出：「德性是先做一個一個簡單的行爲，而後形成的。這和技藝的獲得一樣。當我們學習一種技藝時，我們願意去做這種技藝，於是去做。就由於這樣去做而學成了一種技藝。我們由於從事建築而變成建築師，由於奏豎琴而變成豎琴演奏者。同樣，由於實行公正而變成公正的人，由於實行節制和勇敢而變成爲節制和勇敢的人。」（王海明，2001）[2]

近代演化生物學中談論道德的方向，與亞理斯多德的觀點大異其趣。其談論道德的方向主要著力於：道德感或道德情緒的先天性及其演化（e.g, Alexander, 1987; Axelord, 1984; Boorman & Levitt, 1980; Hamilton, 1964; Trivers, 1971）。不只生物學者，經濟學者也加入了這場道德先天演化論的建構與辯論。雖然大多數經濟學者主張人性是自利的，但是藉著演化賽局的推導，少數經濟學竟也建構出顛覆傳統「理性選擇」（極大化自利）理論的道德先天論。這其中的蹼楚者就是經濟學家法蘭克・羅伯特（Frank Robert）。本文的重點在於引述法蘭克的道德情緒理論（Frank, 1988），由強調人先天本性的演化模型爲出發點，去探討「增强人的內在德性」的道德教育在社會發展上的功能與角色，並討論內在德性與外在規範的關連性的道德教育問題。本節以下將簡要說明道德先天演化理論與法蘭克道德情緒理論的來龍去脈，並說明道德情緒理論與本文所探討的相關重點的承續與關連。

雖然經濟學家常以人性的自利爲分析的起點，然而最早對人性的自利行爲建立理論基礎的反而是演化論。個體行爲的理性自利假設可以遠溯自達爾文（Charles Darwin）在 1859 年發表的《物種起源》（The Origin of Species）。個體因理性自利而使自己的

[1] Virtue 一字在大部的參考資料中，不是譯為「德行」就是「美德」。在本文中此二譯名通用。

[2] 引自王海明 (2001)所引述之周輔成所編《西方倫理學名著選輯》上卷，商務印書館 1954 年版，頁 292。

基因廣泛流佈，而自利的對象是個體本身。在達爾文時代（Darwin, 1859）確立的「自利的對象是個體本身」的假設，隨著生物學基因理論的建立而有了調整。道金斯（Dawkins, 1976）的《自私的基因》（The Selfish Gene）闡釋，不論天擇或性擇，演化的基本自利行爲單位是「基因」，而非生物個體或族群，生物僅是基因的製造者與載體，目的是延續和散播基因本身。順著這種思維，演化論開始產生了一些對利他行爲的解釋，而很重要的一個理論即是漢密爾頓（Hamilton, 1964, 1971）的近親選擇（Kin selection）理論。

近親選擇理論是在解釋家族內部對親人利他行爲或自我犧牲行爲的原因。漢密爾頓認爲：一個基因常能夠藉由製造其載體（個體）的犧牲促進它自己（基因本身）的遺傳流佈，而親代對子代或近親的利他行爲如此普遍存在的理由是因爲：至親大多共有同樣的基因。如果一個個體爲了拯救十個近親而犧牲，操縱個體對親屬表現利他行爲的基因可能因此失去一個拷貝的機會，但同一基因卻因拯救十個近親而得到大量拷貝與保存。威爾森（Wilson, 1975）稱這種族群內部對親人的自我犧牲爲，「硬性利他」（Hard-core Altruism）。近親選擇理論雖然解釋了部份的利他行爲，但卻無法解釋非近親者之間的利他行爲，群體選擇論（Group Selection）與互惠性利他（Reciprocal Altruism）理論則試圖彌補此一學術缺口，不過各有其弱點。

群體選擇論認爲個體的選擇有「利群」的偏好天性，也就是說自利是以群體爲單位。如此一來，如果以利群爲基礎的群體因利益會大於以個體爲自利基礎的群體，則它便會因生存的優勢而勝過以利己（個體爲單位）爲基礎的群體（Wilson, 1975, Ch5.）。然而演化學者 Trivers（1985）提出了所謂「群體選擇的謬論」（The Group Selection Fallacy）指出了群體選擇論的弱點。Trivers 認爲：當個人間的利群程度不完全一致時，少數的***較***自私的利群者，會比其他利群者有更佳的生存機會與繁衍更多的後代，如此循環演化的結果，一個原本以「利群」個體所組成的群體最後將完全質變爲充滿著自私的個體。

經濟學家基於重複性賽局假設下所提出的互惠性利他理論是另一個有關非近親者之間利他行爲的解釋。互惠性利他理論認爲：生物個體之所以不惜降低自己的生存競爭力幫助另一個與已毫無血緣關係的個體，是因爲它們期待日後得到回報，以獲取更大的收益。從這個意義上說，「互惠性利他」類似***投資行爲***，利他是一種變相自利的投資期待。互惠性利他被 Wilson（1975）稱爲「軟性利他」（Soft-core Altruism）。然而「互惠性利他」其實終究是還是自利模型的一種變型，並非真正的利他道德行爲，

這並非本文要討論的範圍。

生物學家 Trivers（1985）提出中介性情緒（Mediating Emotions）（如同情心、憐憫心、報復心、罪惡感等）可以幫助我們解釋很多互惠性的利益期待所無法解釋的現象。例如，人們爲什麼在一個不可能再去的遙遠地方的餐廳用完餐後還留下了小費？互惠性的利益期待顯然不易解釋，而慷慨心或同情心可以提供一個合理的行爲動機。又如，匿名的慈善捐獻行爲顯然不是出於互惠性的利益期待，而同情心、憐憫心或罪惡感等***心理上***的回報或處罰則可以提供較合理的解釋。

雖然間接性情緒可以幫助並合理地解釋非親屬間的「硬性利他」行爲，但法蘭克（Frank, 1988）也指出了 Trivers 理論中的困境。Trivers 理論的困境在於無法証明，這些情緒如何能對個人（利他者）在***物質利益的獲得***上有所幫助，而且其所得要能夠大於（至少不小於）沒有此類情緒的自利者。假如利他者因道德情緒而有利他行爲後的物質利益小於自利者，根據演化的原則：***物質資源較少者子孫將較少***，則在繁衍子孫的競爭中有此道德情緒的個體如何競爭得過極大化自利的個體呢？或至少不被完全淘汰呢？如果有此道德情緒的個體之物質資源較少而且子孫較少，那麼現在的世界就不應存在道德情緒者的基因，而將全是極大化自利者的天下了。

Trivers 理論的困境被法蘭克的道德情緒理論所彌補了。法蘭克在其名著 *Passions within Reason*（Frank, 1988）中，結合了心理學、演化論與賽局理論，建立了一個道德情緒理論。說明在「承諾困境」與「名譽現象」這兩個常見的人際互動情況下，有道德情緒的人的物質報酬比沒有道德情緒的人（完全的自利者或互惠性利他者）要來得大，因此道德情緒的行爲基因在演化中佔有一席之地。

法蘭克的理論雖然屬於道德先天演化論的範籌，但本文認爲可延申至道德教育相關問題的討論。首先，在「承諾模型」中，以合作（道德行爲）與背叛（非道德行爲）爲策略，法蘭克推導演化賽局得出結論：道德情緒不只可以單單從演化中得出（即可以不需由教養而來），而且有道德者與無道德者皆會以一定之比例（演化的均衡點）並存於一個社會中。此一特定之人數比例受行爲（合作與背叛）的報酬結構與辨視（欺騙與僞裝）的成本所影響。然而法蘭克的承諾模型卻也留下另一個困惑。如果人類縱使在沒有道德教化之下也會有道德情緒存在，那麼各個文明社會中的道德教育（增强內在德性的道德教育）的意義何在？我們還需要道德教育嗎？而且在其模型中，在高道德水準的社會裡壞人的得利反而比好人的得利還高，那麼爲何我們要把人教育的更有道德呢？本文認爲道德的先天演化與後天的教養並非互相矛盾的，縱使在法蘭克的

先天論成立的情況下，道德教育仍有其存在與推廣的價值。本文將延續法蘭克的演化賽局模型來探討上述問題。本文第二節將扼要的介紹其法蘭克道德情緒理論中的「承諾模型」，而第三節將運用法蘭克的承諾模型來分析與解釋「道德教育的角色及其對社會的影響」。

其次，在「名譽模型」中，法蘭克籍由配合法則（Matching Law）詳細的說明了道德情緒如何能抗拒眼前的誘惑，有道德情緒者因此在長期中得以建立誠實的名譽，而可能獲得比不能抗拒眼前誘惑的無道德情緒者更大的物質回報。這種「名譽現象」使得道德情緒的行爲基因，亦得以在演化的機制中有機會流傳下來。本文認爲法蘭克的名譽模型似乎隱含著對規範倫理與德性倫理的連結。本文將試圖運用名譽模型說明規範性道德教育在***某些情境***之下，是可以增進人的內在德性。本文第四節將扼要的介紹法蘭克的「名譽模型」，而第五節將運用名譽模型來討論內在德性與外在規範二者的關連性，並且用「名譽現象」的增強來修正第三節中「道德情緒的承諾模型對道德教育的啓示」的部份結論。第六節總結法蘭克的兩個模型在道德教育理論上的參考價值。

貳、法蘭克的承諾模型

一、非理性的行爲與捆綁的承諾

經濟學者對道德情緒的討論由來已久，亞當斯密一生中寫過兩部重要著作，一是「國富論」，另一部是則是「道德情緒論」（常被譯爲「道德情操論」）。亞當斯密的道德情緒論中的「道德情緒」（Moral Sentiments）[3]可被歸類於本文所稱的「不完全理性選擇的假設」（Uncompleted Rational Choice Hypothesis）之下。不完全理性選擇假設的定義是相對於經濟學中的「理性選擇假設」（Rational Choice Hypothesis）而來的，在此有必要先澄清一下，以免讀者在閱讀本文時形成誤解。

經濟學中所謂的「理性選擇假設」，對人性有兩個最基本假設：（1）人性是自利

[3] 根據 Evans (2001/2005) 的解釋，sentiments 這個字，在現代英文中已很少見了，而它的形容詞 sentimental (多愁善感的)甚至還略帶負面的含意。但是在兩百五十年前，sentiments 這個字和現在英文中的 emotions (情感、情緒)的意思是差不多的。在本文中，道德情緒是指亞當斯密所定義的 moral sentiments。

的（Self-interest）；（2）人在作選擇時企圖極大化（Maximizes）其利益。而某些特定的情緒，例如恥辱（Contempt）、憎惡（Disgust）、忌妒（Envy）、羞愧（Shame）、和罪惡感（Guilt）等，被亞當斯密稱爲「道德情緒」，這類情緒能使人對抗那些源自於物質報酬的理性計算，使人***不企圖***完全極大化其利益。這個「不企圖完全極大化自己利益的行爲」，即被本文定義爲「不完全理性選擇」。在下文中提及的「有道德情緒者」、「不完全理性者」、「誠實者」或「合作者」等名詞，皆是在指涉「***不企圖***完全極大化自己利益的人」[4]。

根據法蘭克（Frank, 1988）的看法，我們有很多行爲雖然明知其結果將是不完全理性的，但是仍會去執行它。在此「不完全理性的行爲」意指，人們知道如果不去執行它，其結果將比去執行它得到較好的物質報酬。多數的文獻認爲強烈的情緒和其所引起的非理性行爲會干擾人們追求自利的極大化。故他們認爲激情（不完全理性行爲的動機）最好是被控制著，人們要儘量避免情緒用事。法蘭克的主張剛好相反，激情（Passion）常對利益的獲得非常有幫助。這個看似矛盾的主張並非起因於在激情***演出***下的不完全理性行爲中隱藏著不爲他人所知的附加利益，而是起因於很多重要的問題根本不能被理性行爲所圓滿解決。這類問題有一個共通的特徵，即在某些情境底下，如果我們要解決它，我們必須先作一個違反我們自我利益的承諾。此即是所謂的承諾的問題（Commitment problem）。

承諾的問題起因於，當一個人的利益決定於一個「捆綁的承諾」（Binding Commitment）（例如，如果你這樣做我就***一定***會不計代價地那麼做），而此一「捆綁的承諾」的內容卻是違背個人的極大化利益。亦即承諾的問題是，在某些人際互動的情境中，做出一個違背個人極大化利益的「捆綁的承諾」，其結果反而會比追求極大化自利來的更爲有利。但是理性的人（追求極大化自利的人）在此碰到了困境，他做不出這種承諾，縱使說出了也不易取信於人。法蘭克在他的書中舉很多承諾困境的例子，以下是其中兩個。

（一）合作（Cooperation）的例子

[4] 「理性」一詞在不同的學科中可能有不盡相同之涵義(例如，在康德哲學中與經濟學中，理性的定義可能有很大的差異)，而人們日常生活中使用「理性」一詞亦可能有歧義。因為本文是使用經濟學者的模型為分析工具，故延用經濟學對「理性」一詞的定義。

甲和乙兩人想合開一家餐廳。甲是一個非常有天分的廚師，但卻無管理能力；剛好相反地，乙不會煎蛋，但是個精明的經理人。所以兩人合夥開餐廳會比各自打天下賺得多。但是他們也都知道對方能有機會去欺騙自己而不會被發覺，乙可以從零錢抽屜裡偷走一些零錢而不會讓甲發現；甲可以從食物供應中拿取回扣。如果只有其中一人欺騙對方，則這個人便會賺得比較多，另一人就賺得比較少；如果兩個人都欺騙對方，結果會比兩人都誠實經營生意來得糟。如果甲乙兩人可以約定互不欺騙（即做出一個違背極大化自利的捆綁的承諾），則兩人都可因此受惠。但是如果他們都是理性的人，必然會爲了追求自利的極大化而去欺騙對方，結果會比遵守非極大化自利的承諾來得差。理性的人在此碰到了困境，因爲在承諾的困境中，理性的人很難做出令人「可信的承諾」（Credible Promise）。

（二）制止侵犯（Deterrence）的例子

假設史密斯種大麥，瓊斯在鄰近的土地上養牛。瓊斯有責任爲他的牛隻對史密斯的大麥所造成的傷害負責，他可以建藩籬來預防，但這要花費他 $ 200。如果他沒有建藩籬，牛隻會越界跑去吃掉史密斯價值 $ 1000 的大麥，而史密斯需花 $ 3000 的訴訟成本去控告瓊斯才能取得 $ 1000 的賠償。如果史密斯能做出一個可信的捆綁的承諾（不論花多少錢他都會去法院控告），史密斯的問題就解決了。因爲知道被控告是無可避免，瓊斯瞭解無法從置之不理（不建藩籬）中得到利益，他因此會建藩籬來預防牛隻越界，而史密斯也因此不但可省下訴訟費，而且還可保住大麥不被瓊斯的牛傷害。但是，如果瓊斯相信史密斯是個理性的人，他就不用擔心史密斯會去告他，因爲史密斯將得不償失。理性的人在此又碰到了困境，因爲在承諾的困境中，理性的人很難做出「可信的威脅」（Credible Threat）。

除了上述兩個例子外，法蘭克在書中還舉了，囚犯困境（Prisoner’s Dilemma）、最後通碟的討價還價（The Ultimatum Bargaining Game）、綁架（Kidnapper）、結婚與機會性離棄（另一半是否在有更好的第三者出現時會離棄自己）……等人際互動中常見的承諾問題。法蘭克認爲現今已被提出來的建議常常是含糊或不易被執行的。例如在合作的例子中，當一個欺騙行爲不會被察覺時，我們即很難去相信或做出一個不欺騙對方的承諾。又如在制止侵犯的例子中，當被威脅者知道威脅者的報復是非理性的（即報復的成本大於報復的利益），被威脅者很難去相信威脅者會去實踐這個「將會進一步損害自己利益的承諾（威脅）」。

在文明社會中，最常見而可行的解決承諾問題的建議，是用法律或管理規範去改變物質報酬上的誘因。然而改變物質報酬的誘因並非唯一的解決承諾問題的途徑，在漫長演化的歲月中，遠古的人類祖先是如何在改變物質報酬的誘因的文明技巧未被廣泛運用之前，演化出解決承諾問題的「策略」（Devices）呢？法蘭克認爲在演化中另外還有一種潛在的解決途徑，那就是由道德性緒來產生並使人相信一個不完全理性的承諾。

二、道德情緒扮演解決承諾問題的策略性角色

法蘭克認爲「道德情緒」在「演化」的第一個意義爲，道德情緒扮演著一個解決承諾問題的策略性設計（Commitment Devices）。在承諾問題中最困難的地方就是在創造一個「可信的威脅」或「可信的承諾」。由於這類的威脅或承諾會使自己的利益**非**極大化，理性的人（追求極大化自利者）不易真的做得出來，縱使他真的願意做出這種非理性的威脅或承諾，只要別人仍相信他是一個理性的人，則這種威脅或承諾也變得不可信。理性的人創造不出的這種可信的威脅或承諾，不完全理性的人卻可以創造出來而且能被相信。不完全理性的人之所以是不完全理性的，因爲他有道德情緒。有道德情緒者可能會做出損害自己極大化利益的事，而且別人也傾向相信他會做出這種事情，只因爲情緒的作用非常強大。

法蘭克認爲行爲受兩種心理報酬機制的引導，一方面是理性的物質報酬計算而另一方面是心理情緒的趨力。進食的系統提供了一個很清楚的例子。任何個體進食並不是出自於對熱量需求的理性計算而是生物性「飢餓感」趨力下的結果。相似的觀點可以解釋什麼原因促使我們尋找食物。當食物稀少時，找到食物需要很大的努力，尤其是對因缺少營養而虛弱的人來說更是如此。對這種人而言，強烈的飢餓感比理性的計算更能激發尋找食物的求生行動。

「由內在心理情緒所引發的行爲趨力」與「由理性的計算所引發的行爲趨力」，二者並非總是配合得很好。當一個演化出內在感覺所賴以依存的環境改變時，二者的衝突也可能隨之而起。進食的報酬系統再次地提供一個很好的範例。食物短缺在演化的歷史上被認爲是經常發生的，在此情境下，一旦食物能被獲得則人們會大量進食（吃到飽），人類有動機把自己養胖以作爲對抗飢荒的一種避險手段。然而在當代工業社會中，如此做法將使自己更易於死於心臟病甚於飢餓。在當代社會中，理性的計算產生的行爲趨力可能是「保持苗條」，不管是爲了健康或流行的審美觀。在當代，這兩

種趨力常是處於衝突狀態中。試想像一下，當某人因種種理由，在理性思考之下，下定決心要開始節食了，方法是「餐餐吃半飽」。當在用餐吃到半飽之前，「保持苗條」的理性行爲趨力與遠古演化出來的「儘量進食」的趨力是一致。但是在過了半飽之後，這兩種趨力處於交戰狀態，節食者陷入吃與不吃的掙扎中。相信嗎？理性計算的趨力經常是輸家。想一想爲什麼有那麼多減重瘦身中心和減肥藥廠的存在，就可以證明此點了。

類似的情況也發生在承諾的困境中，有道德情緒者之所以能做出一個違背個人物質利益極大化的捆綁的承諾，乃是因爲演化出來的行爲趨力（道德情緒）常能對抗那些源自於物質報酬的理性計算。由於能做出令人相信的（真的）「捆綁的承諾」，有道德情緒者的物質報酬反而會比追求極大化的理性自利者來得大，因此懷有道德情緒的人類遠祖的物質資源不會比較少，根據演化的原則（物質資源較少者子孫將較少），因此子孫也不會比較少，所以此情緒的基因得以有機會在演化的競爭中流佈。

三、傳訊的問題

至此法蘭克的承諾模型成功的解決了前述 Trivers 間接性情緒理論的困境。不過問題還沒有結束，法蘭克在其模型中還繼續追問：如果道德情緒是如此有用的，爲什麼世界上還會有這麼多不誠實（不道德）的人存在？爲什麼他們沒有被完全淘汰呢？承諾模型可以解釋這類的問題，不過在這之前先要了解的一個重點是，道德情緒雖能幫助我們解決承諾的問題，但這是有傳訊（Signaling）方面的前題的，亦即：個體要有能力發出訊號讓他人清楚地察覺到他的情緒，而且也要能正確地辨認出別人釋放出的情緒訊號。

以前述「制止侵犯」爲例子，道德情緒者獲致比理性的自利者更大之物質報酬的前題至少有二：（1）別人要能透過某種線索與徵兆認得出史密斯是一個在憤恨不平時無論花多少錢也會去法院控告到底的人。（2）而且史密斯也要有能力傳達出某些線索與徵兆讓他人清楚地察覺到他的這種行爲傾向（Behavioral Predisposition）。試想，懷有報復的情緒傾向但不能被他人察知（認出）的人，將很難嚇退潛在的掠奪者，而如果真的不計代價在事後再尋求報復，其後果可能適得其反（利益反而更小，子孫更少）。再以「合作」開餐廳爲例子，誠實者的潛在報酬是在於有能與其他誠實者合作的機會。要使得誠實者能獲得潛在的物質報酬的前提也是上述兩項：（1）別人（其他誠實者）要能透過某種線索與徵兆清楚的察覺出他是誠實者。（2）他也要有能力藉由

某些線索與徵兆認出別人（其他誠實者）亦是誠實的人。認不出來的代價可能是誠實者找到騙子合作而受騙，當然也就不會有較大的利益產生了。

四、欺騙的問題

法蘭克認爲這些傳訊方面的前提是成立（Frank, 1988, Ch.3 & Ch.5），因爲傳訊或偵察道德情緒能力較弱的個體，在演化的競爭中不易有子孫流佈其基因。但隨著這個傳訊問題的討論，也使承諾模型必需面對在承諾問題中必然存在的另一個問題：如果無道德情緒者（騙子）能僞裝成有道德情緒者（誠實者），而獲得更大的利益，那要怎麼辦？這個世界將會是騙子的天下嗎？這是法蘭克模型中所面臨的「欺騙的問題」。

法蘭克用合作的例子來說明欺騙的問題。爲了解釋方便起見，他把模型中兩種可能的行爲策略（有道德與無道德的行爲）簡化，假設在人口中存在兩類人：合作者（Cooperators）與背叛者（Defectors）[5]。合作者在此例中爲有道德情緒者，爲誠實者，他們能克制自己在沒有可能被發現的情形下也不去欺騙別人。背叛者在此例中則是缺乏道德情緒的極大化自利的人，爲欺騙者，他們是投機與僞善的騙子。當這兩類人相遇互動時（例如合作開餐廳），他們所得之物質報酬以分數計算如表一所示。表一中的報酬結構是賽局理論中典型的囚犯困境（亦即承諾的困境）的例子。當這兩類人被丟到生存競爭的環境時，每一類型人的人數演化會視模型中傳訊方面的假設而定，即視此兩類人有多容易能被辨視而定。我們將依次思考下述四種傳訊方面的假設之下的演化「模擬」情況，並以假設 4（最近實際社會情況之假設）之演化結果爲分析「道德教育」的基本模型，進而推導與分析加入道德教育後，社會正義與利益的變化。

表 1　報酬結構

	背叛者	合作者
背叛者	各得 2 分	合作者得 0 分 背叛者得 6 分
合作者	合作者得 0 分 背叛者得 6 分	各得 4 分

[5] 合作者（Cooperators）與背叛者（Defectors）為經濟學與演化賽局中的名詞，常用於標示正反兩類行為。

（一）假設一、合作者和背叛者看起來完全相似

假設合作者和背叛者看起來完全相似。在一個如此假設的人口中，兩類型人將隨機配對。自然地，不論是合作者或背叛者都比較樂意和合作者配對合作，但是事實上他們沒有選擇的餘地，因爲每個人看起來都是一樣的。背叛者和合作者的報酬期望值因此取決於人口中的合作者與背叛者的比率。

舉例來說，如果人口幾乎完全是由合作者組成，例如 99.99%爲合作者。縱使在隨機配對之下，合作者幾乎可以確定會與合作者配對成爲合夥人，因此將得到 4 分的報酬。而人口中非常稀有的背叛者同樣地也幾乎確定會與合作者配對成爲合夥人，因此將得到 6 分的報酬，而此時背叛者的不幸的合夥人（合作者）只得到零分的報酬。在此情況下，合作者的平均報酬爲 4，背叛者的平均報酬爲 6，合作者比背叛者少 2 分。

再例如，如果人口中有一半的合作者與一半的背叛者。 合作者有相等的機會得到零分或 4 分的報酬，平均的報酬是 2 分。而背叛者得到 2 或 6 分報酬的機會相等，因此他們的平均報酬將是 4 分。在此情況下，合作者的平均報酬還是比背叛者少 2 分。

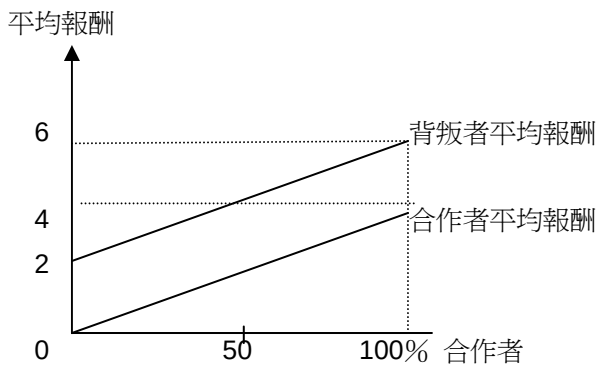

圖 1

事實上，在此傳訊的前提假設下，不論合作者與背叛者的人口比例如何，合作者的平均報酬總是比背叛者少 2 分，如圖 1 所顯示。而根據前述演化論的基本原則：物質資源較少者子孫較少，不論剛開始時的人口比例如何，既然背叛者總是得到較高的報酬，其子孫也總是較多，背叛者的人口比例將隨著時間的經過而成長，合作者的人

口比例將隨著時間的經過而減少，最後整個族群將成爲背叛者的天下，合作者將在演化中滅絕。是以當假設合作者和背叛者看起來完全一樣而不可能被分辨時，真正的合作者（有道德情緒者）是不可能在演化中浮現的。

（二）假設二、當合作者容易被識別時

假設其他條件與前例一樣，但合作者和背叛者彼此是可以**完全並且無成本地**被區別。在此假設下，情況完全翻轉了。合作者現在每次都能選到合作者配對，並且保証有 4 分的報酬。因爲沒有合作者要與背叛者配對，背叛者就只能與背叛者配在一起了，而只得到 2 分的報酬。報酬不再仰賴人口中的合作者與背叛者的比例。在任何人口比例下合作者總是得到 4 分報酬，而背叛者總是得到 2 分報酬（見圖 2 ）。不論剛開始時的人口比例如何，代代相傳的結果，最後整個族群將成爲合作者的天下，背叛者將在演化中滅絕。

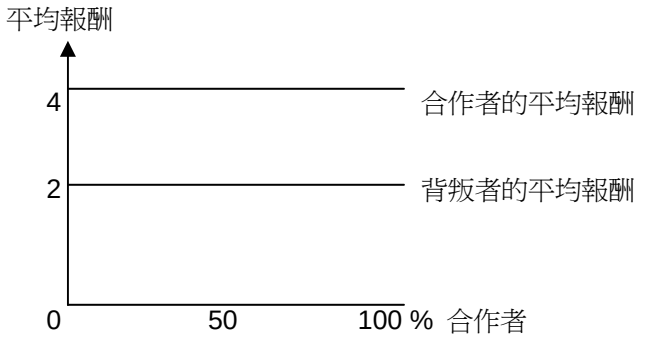

圖 2

（三）假設三、當背叛者可以沒有成本的完全僞裝成合作者時

假設有背叛者發生突變，有些背叛者可以很輕易地在不花任何成本的情況下，僞裝成完全與合作者相同，合作者要區別他們是不可能的。在此假設下，合作者和背叛者看起來完全相似，所以配對的情形又回到假設一的情況了，而結果也與假設一相同。代代相傳的結果，最後整個族群將成爲背叛者的天下，合作者將在演化中滅絕。

（四）假設四、當背叛者無法完全僞裝但辨視其僞裝需要付出成本時

前三個假設都離真實的情況很遠，討論的目的只是爲假設四的說明作準備。現在，假設當背叛者可以僞裝成合作者，但無法完全僞裝，即僞裝有可能被視破，但是辨視其僞裝需要付出成本。爲了簡單化起見，假設報酬還是如表一所示，但合作者辨視背叛者的僞裝需要付出 1 分的辨視成本，如果他付出了，則保証能與另一個合作者配對在一起；但是如果不付出辨視成本則必定無法分辨出誰是背叛者誰是合作者，合作者可視需要決定是否要支付辨視成本。

舉例來說，當原始的合作者的人口比例是 90%時。如果不花辨視成本，一個合作者與另一個合作者配在一起的機會爲 90%，與一個背叛者配在一起的機會只有 10%。他的平均報酬將有（0.9 x 4）+（0.1 x 0）=3.6 分。既然這比花費辨視成本後所得到的 4 – 1 = 3 分報酬更高，顯然在此人口比例下不支付辨視成本比較好。所以此時合作者會選擇不支付辨視成本。

但是，如果原始的合作者的人口比例不是 90%而是 50%時。如果不花費辨視成本，則現在合作者將只有 50%的機會與另一合作者合作。他的平均報酬因此只有 2 分，而花費辨視成本後所得到的報酬爲 3 分，顯然在這種人口比例下，花費辨視成本所得到的報酬反而較高，此時合作者將願意支付辨視成本。

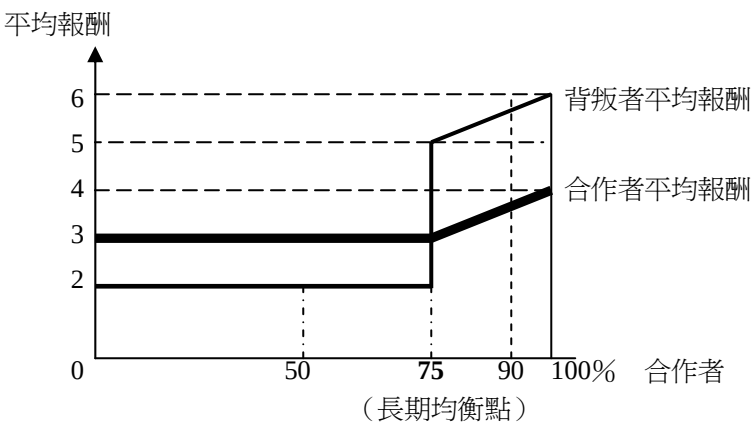

圖 3

事實上，這個例子隱含著一個**長期**「均衡點」（見圖 3），即當合作者的人口比例是 75% 的時候，不支付辨視成本的合作者有 75% 的機會得到 4 分報酬，和 25% 的機會得到 0 分報酬，平均報酬是 3 分，也等於他已經花費辨視成本後的報酬 3 分。即在合作者的人口比例是在 75% 時，花不花費辨視成本，報酬都是一樣的。而**長期**均衡點的位置將隨辨視成本與報酬結構的改變而改變。

圖 3 顯示在各種不同人口比例下合作者與背叛者的平均報酬。當原始的（短期的）人口比例是在 75% （長期均衡點）左邊的時候，合作者將藉由支付辨視成本而總是能找出其他合作者與之配對，得到 3 分的報酬。因爲眼光銳利（付了辨視成本）的合作者不肯與背叛者合作，所以背叛者只好與其他背叛者合作，得到只有 2 分的報酬。因此，如果合作者的人口比例少於 75%時，合作者將會拿到比背叛者更高的平均報酬，意謂他們的子孫比較多，合作者的人口比例將會持續成長至 75%（長期均衡點）。

當原始（短期的）人口比例是大於 75%時，情況剛好相反。合作者沒有道理花費 1 分的辨視成本。但此時背叛者的平均報酬會高於合作者，意謂著背叛者的子孫會較多，合作者的人口比例將會下降至 75%（長期均衡點）。因此在這個例子中我們知道，人口比例將在 75%這個長期均衡點附近來來回回的徘徊，而 75%爲長期的**穩定**均衡點。演化的結果將是合作者與背叛者以 3：1 的比率並存於此社會中。

參、道德情緒的承諾模型對道德教育的啓示

綜合上述合作的例子與承諾模型中的假設四，法蘭克理論可得出三個重要的結論：（1）道德情緒可以由先天的演化中產生，因爲在承諾的困境中，有道德情緒的不完全理性者（合作者）的報酬有可能可以比完全理性的自利者（背叛者）高。（2）在背叛者無法完全僞裝但辨視僞裝需要付出成本時（接近社會真實情況的假設之下），演化的機制將會把合作者和背叛者拉向一個穩定的混合人口比例 （長期均衡點）。道德情緒者（或行爲）與理性自利者（或行爲）可在同一個環境中同時並存下來，誰也不會被完全淘汰。（3）短期的人數比例可能偏離長期均衡點，但長期而言，只要行爲的報酬結構與辨視僞裝的成本不變，人數比例還是會回歸長期均衡點。

法蘭克的模型給了我們上述對道德情緒的演化很有價值的解釋，但同時也留下了一個可進一步追問的問題：假如人類縱使在沒有道德教化之下也會有道德情緒存在，

那麼各個文明社會中的道德教育（增强內在德性的道德教育）的意義何在呢？我們還需要道德教育嗎？又、在圖 3 中，均衡點的右方，背叛者的平均報酬竟會高於合作者的平均報酬，這不免令人困惑，高道德者較吃虧嗎？那爲什麼我們要把人教育的更有道德呢？法蘭克的原著並沒有好好的處理這些涉及道德教育的問題。本文認爲法蘭克的承諾模型本身就育含著這些問題的答案。以下將借用法蘭克上述承諾模型中的合作的例子，並延用表一的報酬結構與圖 3 的人口演化圖形，來分析這個問題。

增强內在德性的道德教育在本文中可被定義爲：「借由增强內在德性，人爲的而且***短期***的改變合作者（好人）與背叛者（壞人）的人數比例，使合作者的人數增加」。在此定義下，可分兩種情況討論。

（一）情況一，假如在某一社會的某一時期，合作者與背叛者的人數比例爲長期的均衡點，延用前述法蘭克的合作的例子，在本例中長期的均衡點爲 75%是合作者，合作者的平均報酬是 3 分，而背叛者平均報酬可能是 2 分也可能是 5 分。假設「增强內在德性」的道德教育使得這個社會中合作者的比例增至 90%，合作者的平均報酬是 3.6 分，而背叛者此時的平均報酬會是（0.9×6）+ （0.1×2）= 5.6 分。在此情況背叛者的平均報酬將確定會大於合作者。

由上述可知，當合作者的人口比例愈多時，背叛者的平均報酬也會愈大。這個情況是可以理解的，即當社會中多數人是誠實者，但他們不願花偵察費用也看不出其所合作的伙伴是否爲騙子，在此的情況下，對少數的騙子而言，遇見誠實者而且欺騙成功的機率增加，平均報酬也會愈大。而根據前述演化論的基本原則：物質資源愈多者子孫較多，顯然在這種社會中，背叛者的子孫不會減少。根據法蘭克的模型，一旦除去（或降低）道德教育，背叛者的人口比例將會慢慢增加，回到長期的均衡點。道德教育只能在短暫的世代中，維持高度的合作者的人口比例，如果不對每一個世代持續進行道德教育，則人口比例終將慢慢回歸 75%的長期均衡點。持續道德教育雖然可使人口比例在長期中不回歸均衡點，但也別想背叛者會消失，因爲當他們的人數愈少時報酬也愈大。

（二）情況二，如果社會本來在短期處於均衡點的左方，而道德教育使社會人口比例在短期內能調整至均衡點的右方。那會是怎樣的情況呢？可分兩個層面來討論：

第一，由各別合作者的利益來看，由圖三可知，在比較沒有道德的社會中（均衡點的左方），個別合作者因背叛者很多，因此會處處提防以免受騙上當，亦即合作者會支付辨視成本，因此其平均報酬是 3 分，而個別背叛者此時的平均報酬是 2 分，背

叛者的平均報酬確定會小於合作者。反之，在一個壞人很少的社會（均衡點的右方），例如前述合作者的比例爲 90%時，合作者不需支付辨視成本，故個別背叛者的平均報酬反而高於個別合作者。

第二，很有趣的，如果以一個社會的總報酬（總產出）來看，情況與個人的遭遇似乎又不一樣。沒有道德的社會的總報酬會較小，有道德的社會的總報酬會較大。以一個社會有 100 人爲例，當社會中合作者的人口比例是 50%時，此社會由人與人合作產生的總報酬爲（50×3）+（50×2）=250 分。而當社會中合作者的人口比例是 90%時，此社會由人與人合作產生的總報酬爲（90×3.6）+（10×5.6）=380 分。

綜合前述情況一與情況二可得下列結論：

1.「增强內在德性」的道德教育使好人的比例增加，但壞人永遠不會完全滅絕，古聖先賢推崇的理想--**大同世界，可以接近但是卻似乎永遠不會達成**。

2. 道德教育使好人的比例增加，而此時各別的壞人的利益是在增加而非減少，亦即不被教化的少數壞人反而比沒有推行道德教育時得利更大。道德教育的推行不會使得壞人會有壞報，更精確的講法是，道德教育的推行使得壞人會有好報。不過我們也不要感到悲觀，因爲重要的是「好人的平均利益卻也是在增加的」，亦即**道德教育會使好人與壞人都得利**。

3. 道德教育的結果不會使好人不吃壞人的虧，相反的，道德教育的推行反而使得好人在遇見壞人時因較不設防（好人多因而不支付辨視成本）而吃虧。而在低度道德的社會，好人因爲會支付辨視成本，好人反而不易受壞人之害。

4. 因道德教育的推行，使一個社會由低度道德進步至高度道德的過程中，雖然個別好人的利益會由「大於個別壞人」反轉成「小於個別壞人」，但是，「社會的總利益是在增加的，而且每個人，不論好人或壞人的利益也都在增加」。 道德教育或許不會帶來「社會正義」（好人有好報，壞人有壞報），因爲壞人不見得有壞報而且壞人的得利在高道德社會中將大於好人，但是***道德教育卻帶來社會利益***。社會的每一個人將因此而變得比以前更富裕。

5. 由前一點似乎也可以解釋，**爲什麼現存的各個文明社會都是自古以來較重視道德的社會？**在族群與族群、社會與社會的競爭中，高道德的族群或社會的總產出將高於低道德的族群或社會。低道德的族群或社會較易在演化競爭的洪流中被淘汰。

6. 最後一點議見是，由前述可知，一個公平正義的社會不是依賴道德教育來達成的。**道德教育帶來的不是社會正義而是社會利益**，一個公平正義的社會的產生可能是

要依靠其他因素，例如：法治、文化、或是政府對社會與經濟制度的干預。

肆、法蘭克的名譽模型

上述道德情緒教育的角色及其對社會的影響是在表一的報酬結構（承諾困境）假設之下的一種簡化的宏觀討論。本節與下一節中將簡介法蘭克的另一個道德情緒的模型 －「名譽模型」，就微觀的行爲層面而言，繼續追問：道德情緒在促使人們放棄追求完全極大化利益的行爲「過程中」，是如何起了作用？並把它延申至道德教育的討論中。

爲了說明名譽模型，法蘭克先舉了一個簡單的例子，用「歸謬証明法」來論証，在一個文明社會中爲什麼人們重視名譽。假設在一個社會中，誠實者和欺騙者各 50 人，誠實者永不欺騙，而且騙子是很理性的，只有在千載難逢的機會（Golden Opportunity）時才行騙，亦即騙子是很謹慎的（Prudent）[6]。假設因騙子是很謹慎的，50 個騙子當中只有 1 人會被抓到。於是這個社會中有 99 人會擁有好名譽。在這種情況下，根據一個人的好名譽而選擇合作伙伴，被欺騙的機率是 49/99，約 1/2。反之，不根據一個人的好名譽而「隨機」選擇合作伙伴，被欺騙的機率是 50/100，也約是 1/2，被欺騙的機率可以說是幾乎相同。在這樣的社會裡，名譽完全沒有意義，社會中的人不應看重他人的名譽，而個人也不須在意自己的名譽。但是文明社會都是注重名譽的社會。如果上述推論過程沒錯，那就是假設出錯了。騙子其實並不是那麼謹慎的（理性的）。

調整一下原先騙子是很謹慎的假設，現在我們假設騙子不是很理性（不是很謹慎）。騙子會在被抓到的機率不小的情況下行騙。假設有一半的騙子最後會被抓到，在這種情況下，一百人中有 75 人會有「好」名譽。根據一個人的好名譽而選擇合作伙伴，被欺騙的機率下降爲 33%（25/75），因此注重名譽是有用的，符合社會的現況。在法蘭克的「歸謬証明」中，我們得出：騙子不是很謹慎的，而社會上的「名譽現象」是合理的。

然而隨之而來的問題是：爲什麼騙子並不是很理性的（不是很謹慎的）？法蘭克

[6] 在此例中，理性者的行為表現是「謹慎」，理性與謹慎在此例中可視為同義。

引用心理學者 Richard Herrnstein 的配合法則（The Matching Law）（Herrnstein, 1970, 1997）來解釋之。

配合法則中的一個重要原則是:「報酬的吸引力（Attractiveness of Reward）」與「延遲實現的時間（Delay）」成反比，而且立即的報酬（或成本）會在心中被***超比例的***大幅放大。例如，實驗中顯示，多數人會偏好選擇在 31 天後獲得 120 美元而非在 28 後獲得 100 美元，但卻會選擇立即獲得 100 美元而非在 3 後獲得 120 美元。換言之，未來的報酬（或成本）在人心中的影響力會被嚴重的***折舊***（discount），而***立即的***報酬（或成本）的影響力常***超比例的***在心中被放大。

在上述騙子不理性的例子中，欺騙的可能利益是眼前立即可得的，被識破的機率雖然不低，但是被識破後可能要付出的成本（壞名譽）是發生在較遠的未來。眼前立即的利益會受配合法則的影響而在**心中**被大幅放大了，騙子因受不了眼前被放大了的利益的誘惑而行騙。

然而配合法則的影響並非不能被突破，人並非總是眼前被大幅放大了的利益的奴隸。相反的，人很明顯的有相當的「自我控制力」（Self-control），去抗拒眼前被放大的利益的誘惑。法蘭克舉了一個相當有趣的例子說：… 就我所知，沒有任何社會的法律會接受一個小偷的抗辦是類似於，『我沒辦法控制自己不去偷眼前可見的那些錢呀！因爲我的行爲是根據配合法則在運作的。』，很明顯的人不能以配合法則來作爲犯罪謝責的藉口。

法蘭克在這裡引出一個問題：爲什麼面對相同的事件，有些人受不了眼前被放大的利益的誘惑而行騙，而有些人可以有較強的自制力去抑制眼前報酬的誘惑呢？「道德情緒」成了法蘭克解釋中不可或缺的主角。法蘭克認爲「道德情緒」在「演化」的第二個意義爲，道德情緒能抑制眼前被放大的報酬的誘惑。回到上述騙子的例子中，如果騙子能不被抓到，則他會得到一個立即的報酬。但是如果他被抓到，則他不只不會得到立即的報酬，反而會有一個壞名譽或處罰。在被抓機率不小的情況下，他的理性評估可能告訴他，這不是一個值得一試的機會，他知道在這種情況下，能抑制自己而避免去欺騙，將能建立好名譽，而在長期中可能將會更有利。但是立即而被放大的報酬仍在眼前誘惑著他呀！反之，有道德情緒者則較可以抑制眼前被放大的報酬的誘惑。

如果行爲者在情緒上傾向於視欺騙爲一種不愉快的感受，亦即他有道德情緒，則他將較能抗拒欺騙的誘惑。如果行爲機制受到被大幅放大的立即報酬所影響，那麼最

簡單的反制那個立即報酬的誘惑的方法便是，當時（**同時**）有另一個感覺（情緒）把行爲往相反的方向拉，在此例中，罪惡感（Guilt）即是這種情緒。重要的是，這種負面的情緒也是在做選擇時就**同時**（立即）發生了，因此配合法則並不能去折舊（縮小）它。配合法則可去折舊（縮小）其他較晚發生的成本，但對立即在心中發生的罪惡感則起不了縮小的作用。如果罪惡感強烈到一定的程度，它可以拒絕這個因欺騙而來的在**心中**被放大了的立即報酬的誘惑。立即的成本才易於抵消立即報酬的誘惑。

由上述得知，道德情緒引起心中立即的成本（罪惡感），進而扮演一個「增強」自我抑制能力的角色。「謹慎」（怕被抓到）在理論上也可能很巧妙的抑制去欺騙的衝動。在那些被抓機會不小的可行騙情況中，騙子理性思考中可能也知道，能抑制去欺騙的衝動將可能得到一個好的名譽，而名譽在未來可帶來利益。然而沒有良心與罪惡感「增強」其自我抑制的能力，縱使騙子「知道」了，也很難去解決因配合法則而來的自我控制的問題。因此，當我們看到一個從未被發現有欺騙行爲（即有誠實的名譽）的人時，我們有理由相信他的此項行爲是（至少部份是）因爲道德情緒（帶來的罪惡感）的因素。

沒有道德情緒的人很難建立起名譽。因爲在配合法則的影響下，僞裝的成本變的很高。沒有道德情緒帶來的罪惡感，理性的騙子很難能抗拒眼前利益的誘惑，而很難在長期建立誠實的名譽，亦即僞善（長期僞善）的難度很高。在此可推知名譽之所以被重視，是因爲它扮演一個「較可靠的」傳訊的角色，傳遞出個人是否有內在的道德情緒。如前所述，行爲「謹慎」與僞善有其困難，名譽因此可以成爲一個辨識道德情緒的可靠線索，本文把以名譽爲辨識道德情緒的線索的傳訊機制稱爲「名譽傳訊（Signaling）現象」。下節中，本文將用此來討論若干道德教育的問題。

伍、道德情緒的名譽模型對道德教育的啓示

一、對前文中「道德情緒的承諾模型對道德教育的啓示」結論的修正

法蘭克道德情緒的名譽模型中的「名譽傳訊現象」可以用來修正與補充前述「道德情緒的承諾模型對道德教育的啓示」的若干結論。如前所述，在名譽模型中，外在名譽成了內在道德情緒的可靠的傳訊特徵，這也是爲什麼文明社會重視名譽，這一點

已被他的歸謬証明所論証了。而***當名譽成了內在道德情緒的可靠的「傳訊」特徵時***，由於好人可以由名譽去辨識別人內在德行的好壞，所以辨視好人壞人的成本應是比原先（第三、四節中）假設的還要低才對。「辨識成本」的降低，將使圖三中的長期均衡點右移，亦即長期而言好人的比例會「自然」增加。例如當辨視成本由原例中的 1 分降爲 0.6 分時，均衡點會由原來的 75%右移至 85%。亦即人口比例中有 85%會是好人，壞人縮減爲 15%（如圖 4 所示）。此時社會的總報酬（總產出）將比長期均衡點在 75%時大[7]。雖然個別壞人的平均得利仍會大於個別好人的平均得利。

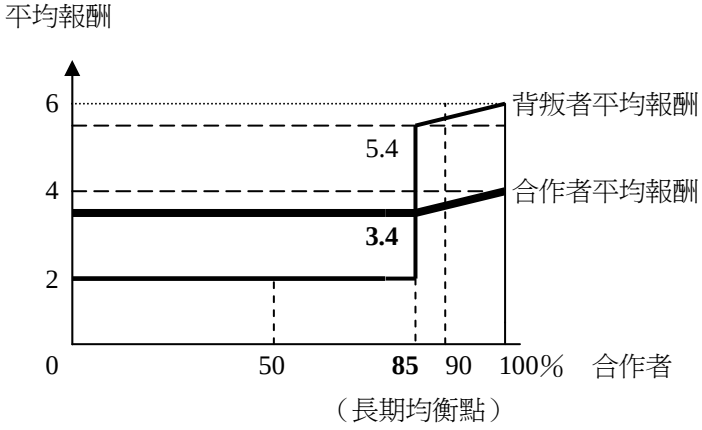

圖 4

雖然本文定義道德教育爲「借由增强內在德性，人爲的而且***短期***的改變合作者（好人）與背叛者（壞人）的人數比例，使合作者的人數增加」，然而由名譽模型可知，增强內在德性（道德情緒）的結果，有助於抗拒眼前利益的誘惑，而在長期建立名譽，名譽因此成了個人內在道德情緒可靠的外在「傳訊」特徵。換言之，道德教育的影響之一是強化名譽傳訊現象，而名譽傳訊現象的增強會使辨識欺騙的成本降低，**長期**[8]而

[7] 以一個社會有 100 人為例，在均衡點為 75%時，總報酬為 (75 x 3)+(25 x 5)=350 分；在均衡點為 85%時，總報酬為 (85 x 3.4)+(15 x 5.4)=370 分。

[8] 請注意，在前文「道德教育的角色及其對社會的影響」一節中，本文是把增强內在德性的道德教育定義為：「借由增强內在德性，人為的而且**短期**的改變合作者(好人)與背叛者(壞人)的人數比例，使合作者的人數增加」。

言道德教育能根本改變演化模型上的均衡人數比例，如上例中，當長期均衡點右移至85%時。亦即道德教育不只有短期的正面影響，也能有長期正面影響。

二、對德行（美德）倫理學可能的參考價值

近年來道德教育的研究有從「責任倫理學」（Duty Ethics）轉而注意「德行倫理學」（Virtue Ethics）的趨勢，而台灣亦有多位學者探討過（黃藿，2003、2002、2000；李琪明，2003；張培倫，2003）。德行（美德）倫理學的重要問題是，「我應該成爲什麼樣的人？」，強調道德教育應該著重行爲者品格（Character）卓越性之培養（張培倫，2003，頁 139-40）。認爲具有一道德人格特質的個人，自然而然容易做出符合道德責任與規範的行爲。然而如何培養人之內在性格？ 此派學說強調道德認知（Moral Knowing）、道德情感（Moral Feeling）、與道德行動（Moral Behavior）。

法蘭的道德情緒的名譽模型似乎可以提供德行倫理學一些參考。首先，兩者皆強調道德情緒，雖然用字不同。法蘭克從演化的觀點認爲道德情緒是先天的，並不直接去討論教育的問題，然而由法蘭克模型中，籍由配合法則詳細的說明了道德情緒如何能抗拒眼前的誘惑，而道德行動才得以成立。情緒與道德行爲（或「內在品格」與「行爲」）的關係在此模型中被很細密的連在一起。近來道德教育學者與倫理學者已在討論把情緒與道德教育放在一起討論（黃藿，2002），但似乎尚未注意到法蘭克的這一有力的論點。

其次，由法蘭克道德情緒的名譽模型延伸來看，罪惡感的增加似乎不見得是一件壞事，這一點也似乎與多數宗教的一些教導原則不謀而合。宗教之所以能增強人的德行（品格），並非只是簡單的天國與地獄的獎懲問題，良心與罪惡感似乎一直是宗教的核心之一。傳統道德教育中的「知恥」與宗教中的罪惡感似乎意外的被法蘭克的道德情緒理論揭發其內在運作的共同性，法蘭克的名譽模型提供了我們思索道德教育的一個可能的方向。再者，近年來教育的思潮比較強調快樂學習與愛的教育（零體罰），但很多人卻感到學生的德行（品格）變差了，當然可能的原因很多，但是爲自己不當行爲所負的責任感、羞恥心或罪惡感的減輕也可能是原因之一。愛的教育與罪惡感的減輕是否有關，可能也是值得我們深思的問題。

最後，由於法蘭克的模型中把內在道德情緒與外在特徵（例如名譽）緊密的論証在一起，外在名譽成了可忠實「傳遞」內在道德情緒的「傳訊」特徵。只要受教者能達成外在名譽的要求，即可估計其內在德行的情況。這似乎提供了教育者一些較可切

實掌握的教育方向，例如軍隊中強調軍人之榮譽心（榮譽是軍人的第二生命），其運作原理即可解釋爲：有名譽者其內在德行應該也不差，因爲在配合法則之下，長期僞善是不易的。看似形式化的榮譽制度，卻隱含對內在道德情緒的挑戰與訓練。由此觀之，某些規範性的道德教育也可以與道德情緒行的外在的傳訊機制（例如榮譽心與同情心）連結在一起。或許這也是研究道德教育的另一思考方向。

陸、結論

本文引介經濟學者法蘭克（Robert H. Frank）在 1988 年所建立的道德情緒（Moral Sentiments）演化理論中的兩個模型，一爲以演化賽局爲基礎的「承諾模型」，另一爲以心理學的配合法則（Matching Law）與自我抑制力爲要素的「名譽模型」，並且進而運用這兩個模型來討論德行（Virtue）方面的道德教育，及道德教育在社會發展上所扮演的角色與價值。而本文的具體重點在於，以法蘭克的道德情緒演化理論爲基礎，加入道德教育因素後，推導「演化模擬模型」，以分析社會正義與利益的變化。並且順便略爲延伸至道德教育的可能途徑。

首先，在「承諾模型」的演化賽局中，法蘭克說明了道德情緒可以只由先天的演化中產生。因爲在承諾的困境中，有道德情緒的不完全理性者的報酬可以比完全理性的自利者更高。在背叛者無法完全僞裝而辨視僞裝需要付出成本時，演化的機制將使合作者和背叛者達到一定的比例 （長期均衡點），道德情緒者（或行爲）與理性自利者（或行爲）可在同一個環境中並存。

在「名譽模型」中，法蘭克則借由配合法則詳細的說明了道德情緒如何能抗拒眼前的誘惑，而道德行動才得以成立。能抗拒眼前利誘的有道德情緒者因此在長期中得以建立誠實的名譽，而可能獲得比不能抗拒眼前誘惑的無道德情緒者更大的物質回報，因此道德情緒的行爲基因得以在演化的機制中有機會流傳下來。

本文首先接續法蘭克的承諾模型論證道德教育在社會發展中的角色與功能。道德教育在此被定義爲：「籍由增强內在德性，人爲的而且**短期**的改變合作者（好人）與背叛者（壞人）的人數比例，使合作者的人數增加」。由承諾模型中的演化賽局推導得出：一個社會由低度道德進步至高度道德的過程中，好人與壞人的平均獲利和社會整體的總利益皆會增加，不過此時好人的平均獲利會由大於壞人反轉成小於壞人。***道***

德教育能帶來社會利益卻不能帶來社會正義。一個公平正義的社會的產生可能是要依靠其他因素，例如：法治或政府的干預。

其次，本文接續名譽模型中的論點推論，在社會有「名譽傳訊現象」時，由於道德教育可以增強內在德性（道德情緒），有助於抗拒眼前利益的誘惑以建立長期的名譽，如此將進一步強化名譽成爲內在道德情緒的傳訊機制，判別好人與壞人的「辨識成本」因此降低。「辨識成本」降低的結果，**長期**而言，將導致承諾模型的演化賽局中合作者（好人）與背叛者（壞人）的人數比例（長期均衡點）將進一步右移，亦即***好人的比例在長期會「自然」增加***。換言之，在長期均衡點上的社會總利益將因「名譽傳訊現象」的增強而增加，雖然此時好人的平均獲利仍然會小於壞人的平均獲利。道德教育不只有短期的正面影響，也能有長期正面影響。

雖然本文一開始設定的討論重點本來不在於探究增強內在德性的***技術***層面。然而本文後來卻發現，法蘭克道德情緒的名譽模型意外地可與道德教育的技術層面做一些連結。因此本文也提供了兩點淺見以供大家思考：

（一）規範層面的榮譽制度與獎懲，如果依賴**立即的**（或近期的）物質報酬或懲罰，學生在遵守規範與否的選擇中，由於「配合法則」的影響，易流於只是追求（或逃避）在心中被放大了的眼前的報酬（或懲罰），如此反而可能訓練出具理性自利的聰明投機者。但是物質報酬或懲罰如果是**遠期的，**輔以增強內在心理的道德情緒（例如同情心、罪惡感、羞恥心），引導其以道德情緒去克制心中被放大了的立即利益，則較可能訓練出具內在德行（品格）的人。是以道德教育技巧層面可做的，可能不只限於宣導利他或行善等此類明顯正面的道德行爲；鼓勵學生把眼光放遠，**認知**到遠期的利益，**追求**遠期的利益，也可能是道德教育的另一個途徑。在追求**遠期的**物質報酬的過程中，人必須自我增強內在的道德情緒，否則不易克服眼前被放大了的立即報酬的誘惑。

（二）由法蘭克的理論可知，道德情緒至少能克制一部份的理性自利的投機行爲。如果道德情緒是情緒的類別之一，則增強一般性的情緒感受力是否也會增強道德情緒呢？一個麻木不仁的人縱使遵守道德規範，我們很難相信他是一位真正具有內在德性（Virtue）的人。但有靈敏情緒感受力而又能遵守道德規範的人，則較有可能是真正具有內在德性的人。例如能由莎翁著作中感受到李爾王的悔悟情緒或哈姆雷特的妒嫉情緒的人，可能自己也較有罪惡感與羞恥心，較易克服眼前被放大了的利益的誘惑。過去談論情緒教育的方向，多著眼於情緒的「控制」、「疏導」或「管理」，但是

如果上述建議成立，則「情緒的深化體驗」亦可能將是道德教育的重點之一。類似於文學方面（包括藝術或電影）的鑑賞力與感受力的培養也可能是德性道德教育的另一途徑。閱讀深入刻劃人性的文學作品，可能可以增強人對情緒的感受力，因而有可能間接增強人的道德情緒與道德行爲。道德教育的內容不限於只在教道德，我們可能可以有不談道德的道德教育。

最後，本文因爲是延續經濟學者 Frank 的模型，所以也延續了經濟學中功利主義的觀點並簡化了問題之複雜性。因此文中作了一些方便於計算成本與報酬的數值假設，而賽局之模擬設定也過於簡單。例如，模擬模型設定爲兩種樣態（即二人）賽局，而非多種樣態（三人或以上）賽局。未來之研究應可多吸收教育學與社會心理學方面的相關資料，令計算之數值假設更接近實際，也可使賽局模擬有更多變化並比較其結果之優劣。

參考文獻

王海明（2001）。利他主義與利己主義辨析。**河南師範大學學報，28**（1），19 - 26。

李琪明（2003）。德行取向之品德教育理論與實踐。**哲學與文化，30**（8），153 -174。

程亮（2004）。道德教育：在規範與德性之間。**湖南師範大學教育科學學報，3**（5），36- 41。

張培倫（2003）。效益與德行：以彌爾效益主義爲例。**哲學與文化，30**（8），139-152。

黃藿（2003）。從德行倫理學看道德動機。**哲學與文化，30**（8），5 -19。

黃藿（2002）。德行倫理學與情緒教育。**社會文化學報，15，**1- 22。

黃藿（2000）。德行倫理學的復興與當代道德教育。**哲學與文化，27**（6），522-531。

Alexander, R. (1987). ***The biololgy of moral systems.*** New York: Aldine.

Axelrod, R. (1984). ***The evolution of cooperation.*** New York: Basic Books.

Boorman, S., & Levitt, P. (1980). ***The genetics of altruism.*** New York: Academic Press.

Darwin, C. (1859). ***The origin of species.*** Cambridge, Mass: Harvard University Press, 1966.

Dawkins, R. (1976). ***The selfish gene.*** New York: Oxford University Press.

Evans, D. (2005). **情感來自演化（*Emotion: The science of sentiment*）**（張勤譯）。台北：左岸文化。（原著 2001 年出版）

Frank, R. (1988). ***Passions within reason.*** New York: W.W. Norton.

Hamilton, W. (1964). The genetical theory of social behavior. ***Journal of Theoretical Biology, 7,*** 1-32.

Hamilton, W. D. (1971). Selection of selfish and altruistic behavior. In J. Eisenberg & W. Dutton (Eds.), ***Man and beast: Comparative social behavior*** (pp. 57-91). Washington, D. C.: Smithsonian Institution Press.

Herrnstein R. (1970). On the law of effect. ***Journal of the Experimental Analysis of Behaviour, 13,*** 242-66.

Herrnstein R. (1997). ***The matching la.*** New York: Russel Sage Foundation.

Trivers, R. (1971). The evolution of reciprocal altruism. **Quarterly Review of Biology, *46,*** 35-57.

Trivers, R. (1985). ***Social evolution.*** Menlo Park, Calif.: Benjamin/Cummings.

Wilson, E. (1975). ***Sociobiology: The new synthesis.*** Cambridge, Mass.: Belknap Press of Harvard University.

投稿收件日：95 年 11 月 1 日
接受日：96 年 1 月 30 日