

版權信息

The Evolution of Morality by Richard Joyce  
 Copyright © 2006 by Massachusetts Institute of Technology  
 This edition published by arrangement with The MIT Press  
 through Bardon-Chinese Media Agency  
 Simplified Chinese translation copyright © 2017 by Yilin Press, Ltd  
 All rights reserved.  
 著作權合同登記號 圖字：10-2012-210號

書  名　道德的演化

作  者　【新西蘭】理查德·喬伊斯

譯  者　劉鵬博 黃素珍

責任編輯　陳銳

出版發行　譯林出版社

ISBN 　　978-7-5447-6980-8

關注我們的微博： @譯林出版社

關注我們的微信：yilinpress

意見反饋：@你好小巴魚

目录

[主編的話 3](#_Toc69137136)

[致謝 5](#_Toc69137137)

[導論　人性 6](#_Toc69137138)

[第一章　助他行為的自然選擇 11](#_Toc69137139)

[1.1　利他主義與利己 11](#_Toc69137140)

[1.2　助他行為的演化過程：親緣選擇 13](#_Toc69137141)

[1.3　助他行為的演化過程：互利共生 14](#_Toc69137142)

[1.4　助他行為的演化過程：直接的互惠 15](#_Toc69137143)

[1.5　助他的演化過程：間接互惠 18](#_Toc69137144)

[1.6　助他的演化過程：群體選擇 19](#_Toc69137145)

[1.7　人類超社會性的演化 22](#_Toc69137146)

[1.8　結論與前瞻 23](#_Toc69137147)

[第二章　道德的本質 24](#_Toc69137148)

[2.1　更新世的愛 24](#_Toc69137149)

[2.2　抑制和禁止 25](#_Toc69137150)

[2.3　何為道德判斷？ 26](#_Toc69137151)

[2.4　道德判斷的實踐影響力 28](#_Toc69137152)

[2.5　道德判斷的主題 30](#_Toc69137153)

[2.6　應得 31](#_Toc69137154)

[2.7　總結與預告 32](#_Toc69137155)

[第三章　道德語言與道德情感 34](#_Toc69137156)

[3.1　動物的規范性（但非道德性）心智生活 34](#_Toc69137157)

[3.2　猩猩沒有道德判斷 35](#_Toc69137158)

[3.3　道德信念的實在主義與工具主義 37](#_Toc69137159)

[3.4　互惠的語言 38](#_Toc69137160)

[3.5　小結 40](#_Toc69137161)

[3.6　情感 41](#_Toc69137162)

[3.7　負罪感 43](#_Toc69137163)

[第四章　道德感 45](#_Toc69137164)

[4.1　預覽 45](#_Toc69137165)

[4.2　作為個人承諾的良心 45](#_Toc69137166)

[4.3　作為人際承諾的良心 48](#_Toc69137167)

[4.4　情感的投射 50](#_Toc69137168)

[4.5　無非是個“原來如此”的故事？ 53](#_Toc69137169)

[4.6　結論 55](#_Toc69137170)

[第五章　用演化維護道德 57](#_Toc69137171)

[5.1　描述的演化倫理學和規范的演化倫理學 57](#_Toc69137172)

[5.2　自然主義謬誤 58](#_Toc69137173)

[5.3　從“是”推出“應當” 61](#_Toc69137174)

[5.4　從演化的角度維護道德：理查茲 62](#_Toc69137175)

[5.5　從演化的角度維護道德：坎貝爾 64](#_Toc69137176)

[5.6　從演化的角度維護道德：丹尼特 65](#_Toc69137177)

[5.7　從演化的角度維護道德：凱斯比爾 67](#_Toc69137178)

[5.8　結論 69](#_Toc69137179)

[第六章　用演化拆穿道德 71](#_Toc69137180)

[6.1　譜系揭穿真相 71](#_Toc69137181)

[6.2　哈曼的挑戰 72](#_Toc69137182)

[6.3　道德自然主義的前景[103] 75](#_Toc69137183)

[6.4　誰需要道德影響力？ 78](#_Toc69137184)

[6.5　重溫哈曼的挑戰 81](#_Toc69137185)

[6.6　可靠主義者、保守主義者、融貫主義者和基礎主義者 82](#_Toc69137186)

[結語　與適應的心靈共存 86](#_Toc69137187)

[參考文獻 90](#_Toc69137188)

[注釋 111](#_Toc69137189)

# 主編的話

劉東

總算不負幾年來的苦心——該為這套書寫篇短序了。

此項翻譯工程的緣起，先要追溯到自己內心的某些變化。雖說越來越慣于鄉間的生活，每天只打一兩通電話，但這種離群索居并不意味著我已修煉到了出家遁世的地步。毋寧說，堅守沉默少語的狀態，倒是為了咬定問題不放，而且在當下的世道中，若還有哪路學說能引我出神，就不能只是玄妙得叫人著魔，還要有助于思入所屬的社群。如此嘈嘈切切鼓蕩難平的心氣，或不免受了世事的惡刺激，不過也恰是這道底線，幫我部分擺脫了中西“精神分裂癥”——至少我可以倚仗著中國文化的本根，去參驗外緣的社會學說了，既然儒學作為一種本真的心向，正是要從對現世生活的終極肯定出發，把人間問題當成全部靈感的源頭。

不寧惟是，這種從人文思入社會的訴求，還同國際學界的發展不期相合。擅長把捉非確定性問題的哲學，看來有點走出自我囿閉的低潮，而這又跟它把焦點對準了社會不無關系。現行通則的加速崩解和相互證偽，使得就算今后仍有普適的基準可言，也要有待于更加透辟的思力，正是在文明的此一根基處，批判的事業又有了用武之地。由此就決定了，盡管同在關注世俗的事務與規則，但跟既定框架內的策論不同，真正體現出人文關懷的社會學說，決不會是醫頭醫腳式的小修小補，而必須以激進亢奮的姿態，去懷疑、顛覆和重估全部的價值預設。有意思的是，也許再沒有哪個時代，會有這么多書生想要煥發制度智慧，這既凸顯了文明的深層危機，又表達了超越的不竭潛力。

于是自然就想到翻譯——把這些制度智慧引進漢語世界來。需要說明的是，盡管此類翻譯向稱嚴肅的學業，無論編者、譯者還是讀者，都會因其理論色彩和語言風格而備嘗艱澀，但該工程卻絕非尋常意義上的“純學術”。此中辯談的話題和學理，將會貼近我們的倫常日用，滲入我們的表象世界，改鑄我們的公民文化，根本不容任何學院人壟斷。同樣，盡管這些選題大多分量厚重，且多為國外學府指定的必讀書，也不必將其標榜為“新經典”。此類方生方成的思想實驗，仍要應付尖刻的批判圍攻，保持著知識創化時的緊張度，尚沒有資格被當成享受保護的“老殘遺產”。所以說白了：除非來此對話者早已功力盡失，這里就只有激活思想的馬刺。

主持此類工程之煩難，足以讓任何聰明人望而卻步，大約也惟有愚鈍如我者，才會在十年苦熬之余再作馮婦。然則晨鐘暮鼓黃卷青燈中，畢竟尚有歷代的高僧暗中相伴，他們和我聲應氣求，不甘心被宿命貶低為人類的亞種，遂把迻譯工作當成了日常功課，要以艱難的咀嚼咬穿文化的籬笆。師法著這些先烈，當初醞釀這套叢書時，我曾在哈佛費正清中心放膽講道：“在作者、編者和讀者間初步形成的這種‘良性循環’景象，作為整個社會多元分化進程的縮影，偏巧正跟我們的國運連在一起，如果我們至少眼下尚無理由否認，今后中國歷史的主要變因之一，仍然在于大陸知識階層的一念之中，那么我們就總還有權想象，在孔老夫子的故鄉，中華民族其實就靠這么寫著讀著，而默默修持著自己的心念，而默默挑戰著自身的極限！”惟愿認同此道者日眾，則華夏一族雖歷經劫難，終不致因我輩而淪為文化小國。

一九九九年六月于京郊溪翁莊

感謝我的母親，

讓我心無旁騖地致力于學術。

# 致謝

該研究計劃最初得到了英國藝術和人文研究理事會（UK Arts and Humanities Research Board）的鼎力資助，謝菲爾德大學哲學系為我的研究也提供了時間豐裕的學術休假。在這段時間里，我的臨時研究助手阿里·馮納埃以其勤勉的工作對我相助良多。學術假期結束之前，澳大利亞國立大學的社會科學研究院為我提供了一份新工作，從而使我得以把大部分精力投入到本書的撰寫中。為此我對邁克爾·史密斯永遠心存感激。此外，我要感謝達特茅斯學院的萊斯利人文學科中心在2004年邀請我去做了半年訪問學者，這里要特別感謝沃爾特·辛諾特—阿姆斯特朗促成了此行。對于參加達特茅斯學院每周閱讀小組的成員，我要特別感謝他們為我提供的反饋和富有啟發性的交流，尤其是歐文·弗拉納根、唐·勒布和錢德拉·斯瑞龐達。回到社會科學研究院之后，我與一群研究生每周會面一次，從頭到尾梳理了本書的最終稿，從中我獲益甚多，為此我要感謝所有相關的人。

對于閱讀整本或部分書稿的人，我都深表謝意，但我在此要特別贊美一個人，即金·斯特爾尼，因為他忍受了我早期書稿的諸多錯謬（在這個過程中，他用完了好幾支紅筆，在書稿的空白處都寫滿批注）。其他給予我珍貴反饋的人包括喬治·博特里爾、布雷特·考爾科特、威廉·凱斯比爾、奧利弗·柯里、泰勒·多格特、里根·法郎茅、本·弗雷澤、丹尼爾·弗里德里克、彼得·戈弗雷—史密斯、吉爾·哈曼、馬克·豪澤、肖恩·尼克爾斯，以及瓦爾特·辛諾—阿姆斯朗。

在我學術假期期間，以下諸位朋友為我提供了安靜的（并且通常還是景色優美的）環境，讓我得以思考和寫作：皮埃爾和瑪麗·鮑利、帕特麗夏·魯賓遜、杰姬和弗里克·阿特金斯、吉姆和凱奧·諾蘭。我還要感謝邁克爾·懷特黑德，他在堪培拉找到了一本弗朗西斯·科布的《道德上的達爾文主義》（Darwinism in Morals）。最后，我要對我的妻子溫迪表達最深切的感激之情，她給予我無盡的愛意和支持。我還要感激剛剛出生不久的馬克斯，他在我撰寫該書第五章之時來到人世，這使撰寫完全書變得尤為困難，但也為我的寫作帶來了更多樂趣。

# 導論　人性

對于善與惡，美與丑，相宜與不相宜，幸福與不幸，恰當與不恰當，什么是我們應該做的，什么是不應該做的等等，來到這世上的人誰沒有一些先天的觀念呢？所以我們每個人都使用這些詞語，而且還力求把我們的先入之見用于各種特定的事例。“某人做得很好，做得不好；做得很正確，做得不正確；某人很不幸，很幸福；很正直，不正直。”我們當中有誰能忍得住不這么說？又有誰要等到學會了這些詞語之后才使用它們？……這是因為，我們來到這個世界之時，自然就已經在這些領域給了我們一定程度的教誨，但從此以后，我們往后又添上了自己的妄自尊大。你問道：“難道我不知道什么是公正，什么是卑鄙嗎？我對此沒有概念嗎？”——有的。“我不是把這個概念用于特定的事例了嗎？”——是的。“那么是我用得不恰當嗎？”——而問題就出在這里，妄自尊大也就在這里開始發酵……現在，既然你認為自己是把先入之見適當地用于特定的事例上，那么請告訴我，你是怎么得出這個結論的？

因為在我看來是這樣的。

可是在別人看來就不是這樣，難道人家不會認為自己是恰當地使用了這些概念嗎？

他會這么認為。

那么，在同一個問題上，有沒有可能你們意見相反，但兩人都正確地使用了各自的先入之見？

不可能。

那么，對于你的先入之見，你能給我們指出一些比在你看來是這樣更好的依據嗎？瘋子不是也只會做在他看來正確的事情嗎？對他來說，這個標準也就足夠了嗎？

不夠。

那么，來找一點比看起來如何如何更可取的東西吧。

——愛比克泰德，“哲學的開端是什么”，《談話錄》

本書試圖完成兩個任務。第一個任務是回答如下問題：“人類道德是先天的（innate）嗎？”（第一至四章）首先，我們需要理解這個問題問的是什么；然后，通過幾個章節的論述，我們會得到一個肯定的回答。由于現有的證據還不足以保證對這個問題做出有把握的肯定或否定回答，因此得出的結論只能是暫時的，并且在一定程度上是推測性的。本書的第二個任務，則是進一步追問：“那又怎么樣呢？”——當然是以一種哲學的口吻（第五和第六章）。如果我們假設道德是先天的（在某種具體闡明的意義上），這能夠以某種方式維護道德，使之避開道德懷疑主義的威脅，乃至鞏固某個版本的道德實在主義嗎？或者，如果道德歸根結底只不過是幫助我們的祖先繁殖后代的東西，這是不是就意味著，道德的權威性反而會遭到削弱，或者用邁克爾·魯斯的話來說（1986：253），“道德是基因強加于我們的各種幻覺的集合”？

對于這本小書來說，這些問題都顯得過于龐大。我在這里不會自稱解決了所有問題；能夠概述一個哲學觀點，為一個正處于發展中的研究計劃添磚加瓦，我就滿足了。我的目標是綜合的與跨學科的，我也知道這種企圖會帶來的危險。但雖然如此，有些迫切而重要的問題在單個學科內不可能得到有效的處理，而本書關注的問題就是其中之一。這些重要問題當然值得一部鴻篇巨著，而我自己開始也是這樣計劃的。但出于一些考慮，我后來認為一本相對較為簡潔和輕巧的書會更符合我的目標，盡管這樣做不可避免地會令一些問題討論得不夠充分。拉爾夫·沃爾多·愛默生曾經指出，“滑過薄冰時，我們的安全就在于速度”，但速度也有其他優點。我希望沒有人會認為，我這在簡短的篇幅里涉及大量材料的雄心，反映了腳下根基的不牢靠。

在能夠評價一個假說的真假之前，我們需要先理解它的內容。說人類道德是先天的，這話是什么意思？首先，“先天”的意義存在爭議。有些理論家論證道，“先天”這個術語過于含混，我們應該把它從嚴肅的學術爭論中剔除出去（Bateson 1991；Griff iths 2002）。我認為這種悲觀的看法缺乏根據，但是我同意，任何在鄭重的討論中使用該術語的人，都應該講清楚他或她的意思。在我看來，人們對“道德的先天性”這個問題進行爭論時，通常指的是我們能否從基因的角度對道德（在某種具體闡明的意義上）進行一種適應性解釋（adaptive explanation）：能否以某種基因型為祖先提供了繁殖優勢來解釋某些現有性狀的存在。[[1]](#_1_171)這并不是說，無論環境如何，一個先天性狀都會發展出來［因為任何表型性狀（phenotypic trait）都不是這樣］，或者它的發展軌跡是被高度規定好了。這也不意味著存在一種“道德基因”。這種先天性概念，以及一般隨之而來的“人類本性”概念，并不蘊涵著任何一種關于人類本質的可疑的形而上學。宣稱直立行走這個特征是先天的，是人類本性的一部分，并不意味著直立行走是成為人類的必要條件（截癱患者聽到這里可以長出一口氣）。

“道德”概念的歧義也增加了理解“人類道德是先天的”這個假設的難度。要澄清這個假設，我們首先需要消除概念上的模棱兩可。一方面，說人類天生是道德動物的，可以是指我們天生就能做出值得道德稱贊的行為——演化的過程使我們合群、友好、仁慈、公正等。任何人，只要停下來環顧四周，就不會說人類的行為總是有德性的，因為我們明顯看到，人類這種生物也可以有暴力傾向、自私、好說謊、麻木不仁，乃至卑劣到無法形容（用馬基雅維利的話來說，人類“忘恩負義、輕浮、虛偽、趨利避害”）。說人類天生就以道德上值得贊揚的方式行動，我們的意思可以是說上述那些道德上令人厭惡的方面是“不自然的”，也可以是這兩方面皆為先天但值得道德稱贊的成分占有主導地位，還可以是說，自然賦予了我們一些值得道德稱贊的方面，跟陰暗面是否存在無關。

另一方面，我們可以用另一種的方式來理解“人類天生是道德動物”這個假設：演化的過程使我們從道德的角度進行思考，生物的自然選擇使我們傾向于使用道德概念。根據上一種解釋，“道德動物”這個術語指的是道德上值得贊揚的動物；而根據這第二種解釋，它指的是進行道德判斷的動物。同前者一樣，后者也可以有多種形式：說我們天生就能夠進行道德判斷，意思可以是，我們被設計成對于某些特定種類的事物具有特定的道德態度（例如，把亂倫和弒父看作道德上可憎的）；也可以是說，我們擁有一種發現某物道德上可憎（或道德上值得贊揚等）的傾向，但憎惡和贊揚的具體內容則取決于偶然的環境和文化因素。這兩種可能代表了一系列立場的兩端；所以，很多中間立場也站得住腳。

以上兩種假設也許有邏輯上的關聯：經常有人論證說，只有那些行為動機是出于道德考慮的人才適合道德評價。刑事法庭界定神經錯亂的傳統標準——即界定被告能否被視為道德上可譴責的標準——是被告能否“辨別是非”（在不少地方，這仍然被看作是判斷神經錯亂的標準）。如果這種關聯是正確的，那么除非人類天生就能夠作出道德判斷，否則就不能認為人類天生是在道德上值得稱贊的；因此，只要能證明第一個假設是正確的，就足以證明第二個假設也是正確的。然而，這個策略看上去并不那么可行，因為這里提到的關聯——這個關聯大意是說，能對某個人進行道德評價就意味著此人的行為是受道德所驅動——仍然存在很大爭議，難以成為有力論證的基礎（下文會提到，其實我懷疑這個觀點是不正確的）。

我在本書關注的是第二個假說，而且我會對其進行直接的考察，而不是通過證明第一個假說。既然點明了我們關注的假說是關于人類做出道德判斷的能力是否先天的，那么就不難看出，從那些支持人類親社會性（prosociality）的論證和材料，并不能必然得出關于先天道德的結論。蜜蜂具有高度的親社會性，但它們幾乎不具備任何道德判斷的能力。但什么才是道德判斷？我們不能回避或輕率處理這個問題，因為除非我們對X是什么已經有了一個充分理解，否則我們對X的討論就會一無所得。這個問題留待第二章處理。

不過在著手處理這些問題之前，為避免誤解，我需要討論一些預備的問題。我們現在談到的觀點（即人類行為或思維在某些方面是先天的），最近幾十年間，在某些人當中很容易激起怒火，以至于光是提起這個問題都難免戰戰兢兢。我確信，對許多讀者來說，我真實的看法非常令人不安，所以我也準備好了面對批評。但要是被莫須有地強加一些愚蠢或有害的觀點，那就實在招人厭煩了。為了避開這類危險，讓我先在某些理論信念和研究計劃的大背景之下，謹慎而清楚地擺明自己的立場。

“生物社會學”（sociobiology）是一項研究綱領的名稱，于1970年初顯雛形。這項綱領以生物學為基礎來解釋社會行為，并預設相關的生物學是先天的。就從生物學角度來解釋蜜蜂、螞蟻或裸鼴鼠的社會行為而言，這項計劃獲得了廣泛的支持，也是成功的。（一份當代生物社會學期刊登出一些無可厚非的標題，例如《論土蝽撫養后代的社會結構》《淡水指甲蛤兄弟姐妹之間的表面競爭》。）只有當這樣的解釋應用到人類行為上，爭議才開始變味。[[2]](#_2_166)不知從何時起，生物社會學研究范圍擴大了，以至于借助生物自然選擇理論來解釋人類行為的方式都被認為屬于“生物社會學解釋”。在20世紀80年代，生物社會學又把自己樹立為演化心理學（evolutionary psychology），但這不僅僅是名稱上的改變。生物社會學關注的是先天行為，而演化心理學關注的是行為背后的心理學機制。兩者的重要區別在于，于史前時期導致適應性行為的心理機制，后來在其他的環境中發生作用時，可能會造成不同的、出人意料的，乃至不適應的行為。這里立刻需要注意三個推論。

（1）演化心理學并沒有聲稱可觀察的人類行為具有適應性，而是認為引起人類行為的心理機制才具有適應性。適應性機制的產物本身不需要具有適應性。

（2）演化心理學也不意味著適應性機制必然造成人類行為跨文化的共性。演化產生的機制可能被設計為對環境敏感的，也可能在預期之外的環境中起到出乎意料的作用，或者在沒有受到恰當觸發的情況下就不能發展出來。只有在這些破壞因素不存在的情況下，我們才能期待有跨文化性出現。盡管如此，如果沒有任何說明普遍的外源解釋因素的更好假說，跨文化性也可以作為先天性的證據。

（3）雖然保留“先天”這個術語最主要是為了說明心理學機制，但是演化心理學似乎也沒有理由談論“先天行為”——這并不是指由先天機制而產生的任何行為，而是一個關于過去的主張：簡略地說，這種行為是由某項先天機制所產生，而且符合這個機制的設計目的（這項機制“理應”在這些條件下產生這種行為），也就是說，這項機制之所以存在，部分的原因就是人類祖先曾經在這些環境條件下進行這種行為。

寬泛來說，沒有任何一個明白事理的人會反對演化心理學。我們到底為什么具有情感？為什么人類記憶力好于金魚？為什么我們能夠更好地記住其他人的面孔而不是數字的順序？為什么我們一般都對性和食物感興趣？為什么我們擁有良好的視力卻看不見紫外線，也不具有回聲定位的能力？毫無疑問，我們必須從人類祖先的生活狀況中尋找答案。因此，有的人很可能會好奇為什么演化心理學引起了如此多的敵意。當中的理由在于，“演化心理學”這個名稱通常所指的觀點，并不只是認為人類心智的基本官能產生于演化的壓力，它還進一步認為，人類心智的許多東西也可以這樣解釋：人類心智包含了大量（可能高達幾百個）先天的、負責特定領域的心理機制，每個機制都是用來應對過去的某些特定的威脅或機遇。但這樣說太過模糊。要認為人類心智在多大程度上可以如此解釋，要贊同多少個特定領域的機制，才能稱得上是（或在某些圈子里被指責為）“演化心理學家”？上面提到的那種有道理而無異議的觀點，和想用生物自然選擇來解釋大部分人類心智的這種導致爭議與分歧的觀點，兩者之間似乎并不存在清楚的界限。毫無疑問，這種模糊性也部分導致了對這門學科的混淆和誤解。

演化心理學承認，大量可以觀察到的人類行為從演化的角度來說是“偶然的”（即這些人類行為是先天機制在新環境的產物），但它同時也承認人類行為的很多方面“理應”具有可塑性——人類心理機制的極強可塑性本身就是適應自然的結果。畢竟，大腦的功能就在于對環境的變化作出多變的反應。因此，我們可以合理地假設，人類非凡的大腦就是被設計出來應對環境變化的杰出器官，而且開放的可塑性正是人類的長項（參見Godfrey-Smith 1996）。人們普遍認為，生物的自然選擇賦予了我們杰出而且功能全面的理性能力。有了這種能力，我們就能對數不盡數的環境刺激應對自如。我們跟動物王國的其他成員的區別似乎就在于，我們有能力“自主地”找到做事情的辦法，有些生存環境雖然迥異于演化設計讓我們適應的環境，我們還是可以成功地應對。我想沒有人會真心懷疑，人類跟所有其他的已知物種不同，人類擁有非凡的可塑性（而我們是否應該把這種可塑性歸于“理性”能力則是另外一個話題）。就想想沖浪（無論海上還是網絡上）、收集棒球卡、思考莎士比亞的著作，或者想想生活在紐約曼哈頓的人們如何成功周旋于日常生活中。面對極為紛繁的環境刺激，我們極為紛繁的反應行為不可能全都是事先“編寫好的”，即便只是使用條件性指令。這是因為，人類祖先的任何生存環境都無法囊括所有類型的環境刺激，再者人類基因組數量有限，無法容納應對這些偶然刺激的所有指令。也許所有這些美妙的多樣性都是“偶然的”（在上文闡述過的意義上），但更有可能的是，人類在多個領域里的可塑性是演化所設計的特征。我們可以合理地認為，面對許多外部刺激，我們的先人依據自己的生活經歷（包括從觀察他人得來的信息）來制定實踐決策，比起依據他/她的祖先過去的經驗來制定決策更為可靠。

另一方面，認為人類心智只不過是具有全面的靈活性的主張，顯然是錯誤的。事實上，沒有任何一個嚴肅的思想家接受徹底的人類“白板說”（tabula rasa）。“白板”這個比喻所指的對象，至少要具有一定獲得、操控和貯備信念的能力——因為人不能通過學習來獲得學習的能力——因此，從環境中獲取信念的能力本身就需要一系列復雜的內在心智機制。甚至約翰·洛克——白板說的創始者——都相信人類具有大量內置的心智能力：記憶、情感、感知、思考、意志、理解力等。對洛克而言，重要的是在人類的白板上沒有刻入任何信念。但甚至連他都清楚，人類白板具有豐富的先天能力與限制。洛克寫道，“我們遇到的所有人，他們品格的十分之九，無論是好是壞，有用沒用，都是教育造成的”［（1693）1989：83］，如果他隱含的意思是人類百分之十的品格是天生的，那么也許有人會覺得這已經足以使洛克有資格成為演化心理學家！

洛克本人無從回答的一個重要問題是：“為什么人類具有這些官能而不是其他？”例如，究竟為什么我們會有情感，而且是恐懼、快樂、嫉妒、內疚等這類情感，而不是其他難以想象的類型？現在，我們可以確定，答案要追溯到人類的過去。有情感，并且是具有這些情感，為我們的祖先帶來了好處，這些特質是被選擇的，我們也由此具有這些特質。當然，在穩定性中也有靈活性。除了大腦受損者，人們都具有恐懼能力，但對什么感到恐懼（豹子、萊姆病、股票市場崩盤，或上帝的震怒），則是從環境中學到的——其中比較重要的是我們從他人那里學到的經驗。不難看出，為什么恐懼這種情感既是固定的，也是靈活的。在人類祖先生活的任何環境中，總會有值得恐懼的事物，但至于究竟是什么事物引起恐懼反應，則會因為不同環境，不同個體，以及個體生命中的不同時段而產生差別。一個很有意思的可能性是，也許有某些確定會引發恐懼的事物，它們以一定的頻率穩定地出現在人類祖先的生存環境中，因此自然選擇的機制發現讓人類固定地害怕某些事物是高效的。[[3]](#_3_161)例如，有人論證過，相比槍支和火藥，人類更容易對蛇和蜘蛛產生恐懼（Öhma et al.2001；Öhman and Mineka 2001）。對于這個個例，盡管還有不少相互沖突的證據，但有一點是肯定的：沒有什么先驗的理由認為有特定目標的恐懼反應不會因此固定下來。另一方面，對于這種精微的適應性，主要的反對因素首先是人類基因組的數量有限，其次是這種性狀也不太可能被單獨選擇出來（也就是說，沒有同時選擇其他有害性狀）。但是毫無疑問，對于一些普遍的、穩定的和重要的環境威脅或機遇，人類具有顯著的固定反應。例如，所有人都喜好食物和性。缺乏食物或性時，我們會更加渴望得到它們。（食物和性同時稀缺的情況下，食物勝出：在救生艇上饑腸轆轆的人們根本不會充滿愛意地凝視著他人豐滿的肢體。）

生物社會學的早期批評者很想貶低這個研究領域，于是不斷叫嚷“基因決定論！”，直到這項指控深入到普羅大眾的心中。但是事實上，任何一個有判斷力的人都不會支持基因決定論，更別說著名的生物社會學家們了。以上討論也足以表明我不是基因決定論的支持者。自然/教育（nature/nurture）的二分法早已經入土為安，乃至現在光是提到它入土為安，都已然令人厭倦。說人類道德在基因層面上是“編寫好的”，意思并不是要否定文化影響的核心地位，甚至也不是說道德的出現完全不可避免。環境因素有多容易對基于基因編碼而形成的性狀的發展造成影響乃至妨礙，這是個實證的問題。例如，苯丙酮酸尿癥（PKU）屬于遺傳性的新陳代謝紊亂，這種疾病可以嚴重地妨礙心智發育的程度，但是單靠飲食上的限制，就很容易避免患病。與之相對，唐氏綜合征是由于遺傳因素而導致的機體紊亂，迄今為止，尚沒有任何方法能避免基因特性的顯現。[[4]](#_4_159)基因型和表現型之間沒有一般性的聯系。因此，史蒂芬·杰伊·古爾德當年的這句話是完全錯誤的：“如果我們是被編寫成為現在這個樣子，那么這些性狀就是不可避免的。我們最多只能引導它們，但是我們不能通過我們的意志、教育或文化來改變它們。”（1977：238）這句話一般而言不可能是正確的，除非他所謂“被編寫”的含義就是不能為意志、教育或文化所改變的東西（這樣的話，他的觀點就顯得沒有什么價值了）。有的人身上的某些性狀雖然在基因層面上是被“編寫好的”，但完全是可以預防的，PKU只是其中的一個明顯的例子。

我們也可以設想，進行道德判斷時的傾向是某種先天的條件性策略的產物。在這種情況下，即使某個社會沒有明確的道德體系，該社會的存在與“道德是人性的一部分”的假設也不沖突。因為這種社會可能沒有滿足相應的前提條件。[[5]](#_5_151)確實，如果我們與我們祖先的生存條件極其不同，那么在原則上，具有道德體系的現代社會就可能不會存在（即可能整個現代社會中都不存在一個有道德的人類），但即便如此，“道德是先天的”這個觀點仍然可以得到辯護。強調這些可能性只是為了說明一點：即便有些東西是人類本質的一部分，它們的顯現也不是必然的。但是我們當然清楚，現代的人類社會具有道德體系；事實上，幾乎所有的人類社會中都有道德體系（參見Roberts 1979；Brown 1991；Rozin et al.1999a）。只有當一個共同體有嚴重而明顯的問題和錯誤時，才會出現類似道德崩潰的狀況。為了找到一個共同體道德完全崩壞的歷史事例，我最初考慮的是歐洲在三十年戰爭[[6]](#_6_145)中所遭受的可怕打擊。在那段時期，紐倫堡的大街小巷里有人肉出售（還有其他駭人聽聞的事件）。但我隨后意識到，出售人肉這個事實表明，即使在那時，某種道德框架依然成立。就算在最可怕的情況下，道德仍然緊緊伴隨著人類——或者更恰當地說，人類仍然堅持著道德。這個令人注目的現象迫切需要一個解釋。

即使道德信念具有先天的基礎，它們也不太可能是無法避免的，這樣一來，就可以駁斥任何類似“用演化論解釋道德在某種意義上剝奪了我們的自由”的擔憂。帕特麗夏·威廉姆斯（1993）曾經論證道，假若道德是先天的，那么我們的道德判斷受到來自內部的強制。然而，要成為有倫理道德的人，我們必須是自由的，由此來看，以演化論解釋我們的倫理在根本上是不融貫的。因此，道德不可能是先天的。對于人類自由意志這個問題，我沒有什么特別要討論的，而且我認為它偏移了現在的爭議焦點。許多哲學家認為，自由并不包括用單純的心靈意志來改變神經因果過程的能力；自由只不過意味著根據自己的欲望而行動。那么，對一個行動的演化解釋也許可以幫助澄清該行動的自由本質，因為它可以解釋與該行動相關的欲望的起源。人類欲望畢竟不是無中生有的——所有欲望都有歷史背景，其中一部分就是演化的歷史。在我們的日常觀念里，信念和欲望都經常受環境因素的激發，但并沒有因此就成為了“被強制的”。我不能明白為什么受基因因素所激發的信念和欲望就有所不同。

人類不同的社群有不同的道德體系，但這并不會削弱道德是先天的這個假說，因為該假說未必就主張具有特定內容的道德根植于人類天性。類似地，說人類擁有先天的語言學習機制，并不意味著日語、意大利語或斯瓦希里語是先天習得的。我們都有能力學習任一種語言，而最終學會的是哪種語言則由社會環境決定。任何道德體系的內容和框架都受到文化的影響，這一點毫無疑問，然而共同體有道德規范這個事實本身可以用某些心理機制解釋，而這些心理機制正是被生物自然選擇所塑造的。盡管如此，自然選擇對道德的內容多少有一點興趣，并且偏愛某些寬泛而全面的普遍內容，這是完全可能的。（我稍后將會討論的證據表明在所有的道德體系中都有些共同的主題。）盡管人類祖先的生活環境多種多樣，這種“固定的”內容可能與某些提高人類適應性的行動和判斷有關。如果環境有變化，靈活性是一件好事；但假若環境的某些因素是保持不變的（例如，我們難以想象在什么持續的情況下，人類吃自己的孩子會提高適應性），具有固定內容的道德態度可能更為高效。畢竟一般來說，表型可塑性（phenotypic plasticity）的代價非常高：要收獲學習的成果，就要冒著反復嘗試的實驗風險，而這可能需要付出高昂的時間代價。（想想在你的皮膚曬黑之前，因為過度暴曬而導致曬傷的難受，或者在免疫系統還沒開始抵抗之前，你因患病而承受風險。）

針對生物社會學和演化心理學，一直有一種抱怨：研究這些領域最終會引發不良的政治后果。謝天謝地，這種聲音近年來有消退的跡象。不要忘了，早期給這些研究領域定調的批評背后有一個政治動機。在《不在我們的基因中》（Not in Our Genes）一書中，理查德·列萬廷、史蒂文·羅斯、利昂·卡姆明等人展開了對生物社會學猛烈而影響巨大的攻擊。他們至少非常坦率地承認，他們都信奉社會主義，還把他們“批判性的科學”看作是“努力創建社會主義社會這個過程的一部分”（1984：ix）。在其他論著中，列萬廷和理查德·列文斯非常驕傲地宣稱：“……我們一直試圖有意識地用馬克思主義哲學作為我們的研究指導，并已獲得一定的成果。”（Levins and Lewontin 1985：165）這種坦承政治動機的行為令人不安，但我在這里要強調的并不是這一點——盡管從智識的角度看他們的觀點確實很討厭（甚至對于馬克思主義者來說也是一樣）。我要強調的是他們的奇特預設：以達爾文主義方法對人類的心理和行為的研究就應該會有一個明確的政治后果。與一些理論家不同，我并不否認達爾文主義的心理學可能有政治含義（因為政治決策原則上可以受任何一類事實信息的影響）。但是，假如達爾文主義的心理學確實有政治含義，那么，要證明這種含義就需要謹慎而復雜的論證，而迄今為止還沒有人提出過這樣的論證。在早期批評言論背后的政治恐懼——害怕生物社會學會為種族主義、性別歧視以及等級主義等開脫——當然是不成熟和缺乏根據的。一直以來都有人憂慮，假若用演化論來解釋一些道德上可憎的行為，這可能就是以某種方式把這些行為變得合理了，或是給那些令人嫌惡的作惡者提供了借口，或是表明了改正這些行為的努力是徒勞的。但是我們已經看到，所謂演化心理學揭示行為的發生是“必然的”，“受基因所決定的”，或者“不是自由的”只不過是稻草人論證（并且一直如此）。[[7]](#_7_143)假設演化心理學揭示了人類大腦的發展是為了使人類心智適應大家族生活。單憑這個假設無法證明這個行為是正當的，也不能證明這個行為是不可避免的，甚至不能證明這種生活安排能夠令我們更幸福。（誰說自然選擇偏愛幸福的生物？）即使演化心理學確實有特定的政治含義，有就有吧。根據政治偏好而接受科學理論，罔顧經驗證據的支撐，僅僅因為實踐后果而否決一個理論，就我目前所知，這樣的決策方式從來沒有資格在人類的知性探索活動中占有一席之地。

人類的心理機制，在多大程度上能用具體而分散的先天官能（而不是用學習過程和功能全面的靈活性）來解釋，對于這個問題，就讓我們的不確定態度同現有的證據保持恰當比例。至于這方面的發現是否具有實踐或道德含義，讓我們假設答案是否定的，除非我們見識到有某個合理清晰的論證支持肯定答案。很難想象會有人反對上述兩個原則。盡管我駁斥了對于達爾文主義心理學的一些錯誤反駁，我還是理解謹慎的必要，并且我也意識到，不是所有的反對觀點都是那么幼稚的。[[8]](#_8_143)在這本書中，我并不認為人類心智的很多方面都可以直接用演化論來加以解釋——對于人類心智，我并不是在鼓吹一個“泛模塊”的觀點，否定全面的組織機制，因此我也不大確定，以下章節所同情的許多觀點是否算得上支持某種（更富爭議性的）演化心理學。對于有關先天道德官能的問題，在我們檢驗證據之前，兩種假說都不應該輕易否決。生物自然選擇應該發展出一些專門用于產生道德判斷的機制，這么說看起來是完全合理的。但與此同時，說自然選擇沒有發展出針對道德判斷的官能，說道德判斷是由文化產生的能力，基于更全面的心理官能，這看起來也很是合理。本書的大部分內容，即前四章，都將致力于處理這個難題。

# 第一章　助他行為的自然選擇

## 1.1　利他主義與利己

道德思維有什么價值？從道德的角度評價自己和別人，能帶來什么好處？我們可以合理地假定，這些問題的答案與幫助他人、相互合作有關。我們都隱隱約約地覺得，一個愿意使用“美德”、“義務”或“正義”這些概念來設想人際關系的人，比起對這些概念完全陌生的人，更有可能是有益的社會成員。本書第四章將詳細分析這種想法，而我現在提及這點，是因為有人可能認為，這個想法對道德先天性的假說構成了直接的阻礙。有人會說，自然選擇不就是一場比賽，最后的冠軍屬于不羈的個人主義者嗎？本書第一章的主要目標就是要指出這個疑慮是錯誤的。我將會概括出幾種方式，來說明幫助他人和相互合作的性狀可能是如何演化出現的。但在開始這個任務之前，我們首先應該弄清楚一些容易被誤解的術語。對于助他行為（helping）的自然起源這個主題，大多數相關文獻都宣稱自己研究的是利他主義行為（altruism）的演化過程，但在多年來對演化論的討論中，利他主義這個術語總是被處理得含糊其辭，敷衍了事，導致某些領域總是混亂不堪，結論也常常錯誤。對于本書來說，在這一點上做到概念精確是非常重要的，因此我下面先做出一個三重區分。

助他行為（helping）能以某種方式為他人帶來好處的行為。與此相對的是有害的行為。［我也樂于把這類行為稱為“合作行為”或“親社會行為”（“prosocial behavior”）。］

犧牲繁殖適應性的行為（fitness sacrif icing）在某種意義上以犧牲自己的繁殖機會為代價，促進他人的繁殖適應性。（這通常被稱為“演化利他主義”。）與此相反的則是增強繁殖適應性行為（f itness advancing）。

利他主義（altruism）目的是給他人帶來好處，行為動機是對他人福利的、非工具性的關心。與此相反的是自私。

這樣限定“利他主義”概念，我覺得是符合日常英語的相關用法的。在英語里，只有當一個行動的審慎動機是以他人利益為導向時，也就是說只有當他人的幸福是行動者的最終理由時，這個行動才是利他的。假設艾米做出了有利于伯特的行為，但是艾米之所以這樣做，是因為她相信這種行為長遠來看會給自己帶來好處。那么，這只是一個自私的行為，不是利他的行為。假設艾米的想法結果是錯誤的，她從來沒有從有利于伯特的行為中得到回報，獲益的人只有伯特一個。這會令我們改變先前認為艾米的行為自私的判斷嗎？不能。一個行為是自私的還是利他的取決于驅使行動者行動的審慎動機（行動者據以而行動的考慮[[9]](#_9_137)）而非取決于誰最終獲益。有人懷疑，從這個角度看，不存在任何利他主義的人類行為；他們認為，所有行為的最終目的都是為了行動者自己的利益。我們接下來將會了解，這些人（心理自我主義者）的觀點幾乎肯定是錯誤的。很多人類行為都真心地出于對他人利益的關心，而非出于自己獲利的考慮。

人類以外的動物，很少（如果有的話）可以毫無異議地被認為具有人類那樣的動機理由（motivating reason，即慎思時的考慮因素）。另外，人類以外的動物具有自我概念的顯然少之又少，但這種概念是形成自私動機的必要條件。所以，人類以外的動物，很少（如果有的話）被認為必然具有利他或自私的動機。我這里并不打算對動機或意圖概念給出精確的定義；對于什么生物具備這些概念，什么生物不具備，我也不想劃分界限。對我來說，只需要指出這一點就足夠了：植物不具有任何類型的動機或意圖。盡管有些植物可能以某種方式有助于其他植物，也可能有犧牲繁殖適應性的行為，但我認為最好還是不要把植物的行為描述為“利他的”或“自私的”。

生物能互相幫助，這當然是再明顯不過的了，但在我們完全確定“幫助”的含義之前，還需要澄清不少概念上的困難。[[10]](#_10_135)一般來說，“讓X獲益”的意義等同于“促進X的利益”，但這就涉及“利益”的含義（就算我們討論的對象是人類，而人類當然有各種利益，事情也不是這么簡單）。如果有人在廚房地板上踩死一只蟑螂，說這“損害了它的利益”也很正常（這樣說并不需要陷入任何可疑的擬人論，不需要認為蟑螂也具備有意識的欲望、動機，乃至還可以體驗到痛苦），但是要確定一只蟑螂的利益是什么，其實沒有那么容易。也許歸根結底，這個概念是無法闡明的，所以要求闡明只是因為糊涂。但重要的是，我們并不是輕率地就把生物的利益等同于它的繁殖適應性。一只動物的利益可能恰巧是它的繁殖適應性（這也可以是我們對蟑螂的利益的分析結果），但情況不必總是這樣。當我為我的朋友購買一件生日禮物，這可以說是一種幫助行為（即便我只是想為我的朋友帶去一點快樂）。但我想要促進的是她（經歷生老病死的這個人）的利益而不是她的基因的利益。比起“蟑螂的利益”，“基因的利益”這個說法更加站不住腳。基因在復制自身的活動中不具備任何利益，正如一條河沒有被大壩攔截住，不能說這符合這條河的利益那樣。然而，一個基因之所以具有某些特性，是因為這些特性可以幫助基因成功復制和延續；因此，我們可以說（用一種半隱喻的方式）基因是“被設計”成為要復制自身的，說復制是它的“目的”。如果我們允許自己這樣說，那么說基因的復制活動對它而言“有利益”，看上去就難以否認了。

如果接受這種說法，認為我朋友的基因和她自己一樣，都擁有各自的利益（或者，也可以說與她“個體的利益”相對的是她“遺傳的利益”），那么作為她的朋友，我挑選生日禮物送給她是出于對哪種利益的考慮，答案是明擺著的。我可能送給她足夠一年用的避孕套作為生日禮物，由此阻止了她的基因復制，但這并不意味著我對她的幫助就減少了。在相反的情況下，我關心的可能是她基因的利益（雖然這聽起來很怪），所以可能暗中破壞送給她的避孕用品。但是，以這種方式促進朋友的基因的利益，并不能說我的行為對她是有幫助的。與此相似，殺父或弒母的行為可能促進了殺人者的基因的利益，從而也就促進死者基因的利益，但顯然并不因此就降低了對于死者的傷害！除此之外，說某人由于被迫建立一個大家庭而損害了自身的利益，也沒有什么不一致的。簡而言之，將一個人與她的基因混淆起來是愚蠢的，這就相當于把一個人等同于她的肺或淋巴結，而一旦我們區分了這兩者，我們也必須區分一個人的利益和她基因的“利益”。

我承認，對于不能體驗到焦慮或痛苦的生物來說，也許可以把它們的利益等同于增進繁殖適應性。不妨做個有用的思想實驗：如果我們發現，一個生物既無法繁殖后代也不能給它的近親在繁殖上提供幫助，捫心自問，我們對于什么會傷害這個生物體的看法，是否會因此而改變。在這種情況下，我們無法影響到它的繁殖適應性；因此，如果我們仍然愿意把踩死或殺死一個生物體看作是“傷害它”（我認為我們經常是愿意的），那么這里一定在使用另一種的“傷害”或“利益”概念，即對于個體的傷害。[[11]](#_11_131)（至于對于該生物的幫助，情況也類似）

至于“犧牲繁殖適應性”的行為，毫無疑問是有的。人們有時候會放棄組建家庭，來到遙遠的國家，把自己奉獻給那里的慈善事業。也有很多人，在戰爭中為了拯救他們的同伴而犧牲自己。如果他們為之犧牲的同伴是他們的家庭成員，那么他們英勇的犧牲行為仍然算是“增強繁殖適應性”的行為。但在很多情況下，他們為之而無私犧牲的人跟他們并沒有血緣關系。就犧牲繁殖適應性的行為而言，事實上引起爭議的問題不是它是否發生，而是這種行為的發生是否受到自然選擇力量的驅動。有些人爭論道，自然選擇不可能選擇犧牲繁殖適應性的行為。例如，理查德·亞歷山大曾經斷言（1987：3），沒有“絲毫的證據”表明，這種類型的行為是“所有人類個體的正常機能的一部分”。從這里引申的推論大概是說，人類之外的動物界不存在任何非偶然的犧牲繁殖適應性行為。

在這一章接下來的論述中，我們會看到，與上述推論相反，犧牲繁殖適應性行為很有可能是通過生物的自然選擇而產生出來的。關鍵在于自然選擇發揮作用的生物群落有一定的群居結構。但事實上，這方面的爭論其實在這里并不特別重要：對于本書的目的而言，重要的是被自然選擇出來的助他行為。至于協助他人的行為有沒有構成犧牲繁殖適應性的行為，或是不是一種促進繁殖適應性的方式，我樂意擱置不管。我感興趣的問題是：“為了控制管理這些助他行為，自然選擇青睞的近因機制是什么？”一個可能正確的答案是“利他主義”。換言之，為了令一個生物順利地幫助他人，自然選擇可能偏向于那些從利他主義動機出發的行為（假定生物具備復雜的認知能力，能夠有行為動機）。另一個我也贊同的答案是“道德”。為了令一個生物順利地幫助他人，自然選擇可能青睞那些做出道德判斷的性狀。本書的主要任務就是探索研究第二個答案。

本章余下內容的任務，是找出可以產生有助他行為生物的演化過程。這可以令我們更好地探討以下問題：道德感是不是一種在人類中發展出來的操控助他行為的手段？但這里有一點先要謹慎對待。假設上述方法是正確的，我們確定人類的道德思維受到特定機制的操控，而這個機制是經過達爾文式的自然選擇過程形成的。可是，要是由此得出這個結論，說所有的人類行動，甚至助他行為和被視作體現美德的行為，“實際上都是自私的”，可就是大錯特錯了。上文做出的區分可以充分表明這個結論錯在何處，可是，這種妄下結論的傾向似乎既普遍，又有害，所以有必要再三強調其錯誤所在。理查德·道金斯基于“自私的基因”的觀點（1981：3），得出“我們天生就是自私的”的結論。亞歷山大（1987：3）寫道，除非我們了解到社會是“個體尋求他們自身利益的集合”，否則我們就無法理解人類的行為。而邁克爾·蓋斯林曾說過一句令人印象深刻的話（1974：247）：“抓破一個利他主義者，看看一個偽君子流血。”但是這種觀點，把自己打扮成強硬的現實主義者，其實錯誤地混淆了不同的解釋層次（參見Tinbergen 1963）。他們尤其犯了一個基本的錯誤，把心靈狀態的原因和它的內容混作一談。一個人對臨近的工作面試感到非常緊張，這部分是因為他剛剛喝了四杯濃咖啡（要是沒有喝咖啡，他也就不會感到這么緊張了），但是下結論說他是對咖啡感到緊張就是胡說八道！認為演化論揭示了我們所有動機、理性和利益的“真正”內容，其實也正是犯了這樣的推理錯誤。假設福瑞德在照顧他患病的妻子。別人問他為什么這樣做，他非常真誠地回答，因為他愛他的妻子，希望可以減輕她的痛苦。而一個演化心理學家可能會告訴我們，福瑞德照顧他的妻子只是為了提高他繁殖后代的機會，因為這樣一來，他的后代就會得到照顧，而他對妻子的感情不過是一套近因機制的產物，自然選擇通過這套機制，使得一個人在配偶需要的時候幫助她。這樣，我們用演化論來解釋了一種認知的/感情的/行為的現象：福瑞德對他妻子的愛。但是，這種解釋無法說明福瑞德的動機內容，也無法證明他“真正”關心的是他的繁殖適應性，他只是間接地關心他妻子的幸福。

有人可能對此提出反對意見，認為這樣的類比不合適。他們可以提出，演化心理學的解釋訴諸的是本身就具有利益偏好的事物（與咖啡不同）：基因。但是，我們已經看到，所謂基因擁有（字面意義上的）利益這種說法是十分不可靠的。（我認為，這個隱喻性用法引起的概念混淆遠遠多于它帶來的理論價值——對于這種概念上的混亂，任何一本討論道德演化的專著的作者在書的第一章都不得不費時間予以回擊。）但即使我們從字面意義上討論利益問題，由此得到的論證仍然是不可靠的。這個論證需要接受以下“利益轉讓原則”：

如果X擁有利益a、b、c，等等，并且X擁有這些利益可以被另一個事實所解釋：Y擁有利益p、q、r，等等，那么，X的利益“從屬于”Y的利益，并且，事實上，X的“真實的”或“最終的”利益是p、q、r，等等。

但是，我們沒有很好的理由去相信這個原則，反倒有好的理由反對它。這個原則繼續混淆了解釋心靈狀態（或利益）的來源與提出心靈狀態（或利益）的內容。這個常見的概念混淆的根源或許是因為“一個理由”有歧義。福瑞德關心妻子的理由在于他妻子正遭受病痛的折磨。這個理由促使他采取行動，也成為他的考慮因素。妻子的痛苦之所以能夠驅使他做出行動，其理由（或者更好地說，其理由之一）則在于，照顧一個人的配偶可以提高他自身的適應性，因此這種行為是自然選擇作用于人類的結果，而福瑞德就是人類。當我們用一個人的基因具有促進復制的特性，來解釋他的行為和心靈狀態，那么，我們就為他之所以具有這樣的心靈狀態和行為提供了理由。但因此就把這些理由視為他的理由（即他據以行動的根據），則是一個嚴重的錯誤。我們同樣可以對雪崩發生的原因產生疑問，但這樣做的時候，我們不會懷疑融雪藏有什么惡意動機。[[12]](#_12_129)我并不是說，一個人的理由對她而言一定總是明顯的；我只是認為，一個人的理由并不總是“歸根結底”跟基因的繁殖有關。

本節所確定的行動的三個范疇，可以任意結合在一起。如果暫時忽略那些宣稱不存在利他行為的悲觀論者，我們可以得到滿足下列條件的例子。[[13]](#_13_125)

——助他的、犧牲繁殖適應性的行為，并且是利他主義的。

——助他的、犧牲繁殖適應性的行為，但是自私的。

——助他的、促進繁殖適應性的行為，并且是利他主義的。

——助他的行為，是促進繁殖適應性的和自私的。

——非助他的行為，但是犧牲繁殖適應性的和利他主義的。

——非助他的行為，并且是犧牲繁殖適應性的和自私的。

——非助他和犧牲繁殖適應性的行為，但是利他主義的。

——非助他的行為，并且犧牲繁殖適應性的和自私的。

有了這些區分，并且已經把概念上的潛在混亂消除在萌芽之中，我們就可以回到論證的第一步。需要注意的一點是，我們關注的并不是利他主義——不管是日常的心理學意義上的利他主義，還是演化論意義上的（我一直稱其為“犧牲繁殖適應性行為”）。利他主義對本書而言不是一個重要議題。我們首要關注的也不是“道德”行為。相反，我們的最初任務（也是本章接下來的任務）是概括出導致助他行為的演化論過程。在合適的時候，我們會討論道德思維可能是人類操控助他行為的機制——但這些都是后話。有關道德的議題是我們的研究目標，但現在暫時還不會涉及。

正如生物之所以能夠移動，有許多演化上的理由，生物之所以能夠彼此幫助，也同樣存在不少演化上的理由。因此，我的目標并不是揭示助他行為被自然選擇出來的唯一方式，而是勾畫幾種大致的演化動力：親緣選擇（kin selection）、互助共生（mutualism）、互惠（reciprocity）以及群體選擇（group selection）。最后，我將會討論文化因素是如何影響人類助他行為的特征。

## 1.2　助他行為的演化過程：親緣選擇

總被率先提及的是被稱作親緣選擇的現象。這方面的經典論述（雖然親緣選擇是達爾文[[14]](#_14_121)隱約有所察覺的一種選擇性動力）是威廉·漢密爾頓1964年的文章《社會行為的遺傳演化》（“The Genetical Evolution of Social Behaviour”）。按照理查德·道金斯著名的建議，我們不妨把生物體設想為基因用來成功復制自身的載體。一個對家庭成員（即那些必定與它分享了基因的生物體）友好并樂于提供幫助的生物，可能就是基因的有用載體。對于基因來說，如果它的“載體”為了拯救三個子女，或三個同胞兄弟姐妹，或九個表親，而犧牲自己的生命，那么這算是一樁劃算的交易。犧牲生命的情況多少有點極端；我們同樣關注更為溫和的犧牲：你把食物分給你的兄弟姐妹，照顧你年幼的侄子（女）或外甥（女），教育你自己的孩子等。生物都應該關心它們自己的子女，這個想法深深地根植于我們的腦海里，因此我們要花一番努力，才能拉開距離，意識到這需要從演化的角度加以解釋。很多生物并不關心它們的后代，它們的繁衍是以數量而非質量取勝。但是，大多數哺乳動物都費心培養高質量的后代，這就要求它們提供一定程度的照顧、食物和教育。人類的嬰兒極大地依賴他人的幫助，并且這種依賴需要持續很久。因此，我們應當預見，人類照顧自己后代的性狀是自然選擇的結果。如果基因的人類載體不愿意照料自己的后代（讓嬰兒一出生就自生自滅），那么這個基因也將很快消亡。

讓我們考慮一下膜翅目的社會性昆蟲，比如螞蟻、蜜蜂和黃蜂。我們可以找出這些昆蟲三個有趣之處。第一，它們是社會生活的成功典范。如果有一種個體組成的群體可以稱得上是“超生物”（super-organism），那就首推螞蟻或蜜蜂的巢穴。第二，它們展示了大量不同尋常的助他行為。蜜蜂擁有以自殺保護蜂房的毒刺。螞蟻當中的工蜂不具備繁殖能力，它們每天的工作就是照料其他螞蟻的后代。演化論需要解釋這些昆蟲社會生活的特殊性。不育和自殺傾向的性狀是如何通過生物自然選擇而形成的？難道自然選擇不是更青睞那些沒有毒刺的蜜蜂嗎？事實上，達爾文已經意識到，社會性昆蟲的特點給他提出了難題，“這難題對我的整個理論初看起來是難以克服的，事實上也是致命的”［（1859）1998：352］。漢密爾頓則通過強調這些昆蟲的第三個特點，提出了不為達爾文所知的答案：它們基因的彼此關聯（genetic interrelatedness）。從基因的角度看，生活在同一個巢穴里的螞蟻、蜜蜂或黃蜂，大多數個體之間的親緣關系比起群居的猴子、土撥鼠或人類更為接近。我們所熟悉的哺乳動物與它最直接的近親成員之間最多只有50%的基因是相同的（除了同卵雙胞胎）。[[15]](#_15_121)膜翅目昆蟲則因為其染色體的特殊排列，呈現出另一種狀況：雄性蜜蜂只有雌性蜜蜂染色體數目的一半，而巢穴的雌性“姐妹”們，作為這個巢穴中占絕大多數的成員，則有高達75%的染色體是一樣的。[[16]](#_16_121)漢密爾頓規則（Hamilton’s Rule）指出，自身付出代價來幫助其他成員的性狀，只要滿足如下條件就會被自然選擇青睞：

r B＞C

這里r是對受惠者和施惠者之間的基因關聯程度，B是受惠者得到的利益，C是施惠者所付出的代價。在膜翅目昆蟲社會中，r經常高于同種哺乳動物的情況，這使得C的數值相應地增大。因此，我們可以預測蜜蜂的犧牲性助他行為的數量要遠遠大于老鼠——這事實上也正是我們觀察到的情況。[[17]](#_17_115)

這樣，對于自然選擇為什么以及如何使生物承受一些代價來幫助近親成員，我們就得到一個非常合理且被高度證實的理論。然而，這看起來似乎無助于我們解釋人類的道德，因為從道德的觀點看（至少在西方的傳統里），偏袒家庭成員的傾向被認為是“裙帶主義”。此外，親緣選擇似乎無法解釋對非親屬的幫助是如何演化出現的，而后者是人類道德的重要內容之一。如果兩個生物個體之間不具備血緣關系，根據漢密爾頓規則，r的數值就為0，r B的數值也因此為0，那么親緣選擇理論就無法解釋任何C＞0的助他行為（即那些無論如何都要付出代價的助他行為）。

然而，對于解釋人為什么會幫助沒有血緣關系的人，親緣選擇仍然是其中一個重要因素。首先，我們應該記住，幫助親屬這一特征必須包括某種近因機制，使其能識別親屬，但這種機制是會出錯的（尤其在一個全新環境中），從而引起對沒有親緣關系的個體的幫助行為。對很多物種而言，氣味是親緣識別的途徑；例如，一個母親或父親產生撫育行為，可能因為新生兒的氣味激發了它們的荷爾蒙（Yamakazi et al.2000）。但是，與此相比，親緣確認（kin identifying）的機制可能更為粗糙。如果組成群落的只是小型家庭，頻繁打交道的同類很有可能與自己具有親緣關系，那么自然選擇就能找到一個簡單的解決方案：“幫助那些你經常往來的同類。”有個很好的例子可以說明自然所采用的粗糙機制：孵化的小雞會“銘印”（imprint）第一眼看到的移動物體，無論該物體是人還是旋轉的紅色立方體（參見Lorenz 1937；Bateson 1966）。在自然環境中，由于小雞剛孵出來第一眼看到的物體通常是它們的母親，因此，這樣的機制可以運作得非常良好。我們在人類社會當中也可以發現類似的現象。人類學家約瑟夫·謝潑德（Joseph Shepher；1971，1983）研究了在基布茲[[18]](#_18_113)社區中被撫養長大的人，他發現，這些人極少與那些跟他們一起長大的人發生性關系，無論他們彼此之間有沒有基因聯系。解釋這種現象的假說（最早由愛德華·韋斯特馬克于19世紀提出）認為，這種機制最初的目的是為了避免親屬亂倫。為了使人類避免與親兄弟姐妹發生亂倫，自然選擇沒有在人類中發展一個“兄弟姐妹探測器”，而是偏向更為簡單的“兒時熟人探測器”。在祖先的環境里，這兩個機制挑選出來的對象幾乎是一樣的，而后一種機制運作成本更低（這個機制在當代得到證實，參見Lieberman et al.2003）。倘若人類幫助親屬（或某些親屬）的行為是由“提供幫助給那些你時常打交道的同類”的機制來調控的，而在當今人類社會中，我們接觸的同類數量遠遠超出自然選擇所設想的（包括從電視、報紙之類提供的“虛擬關系”中接觸到的），那么在其他情況相同的條件下，光是依據親緣選擇，我們就能夠預料到有大量幫助非親屬的行為出現。

對于幫助非親屬的行為而言，親緣選擇之所以重要的第二個原因在于，我們至少可以在親緣選擇中，找到一個對某些生物如何以及為什么具有調控助他行為的機制的解釋。生物自然選擇是一個保守的過程，它把舊的結構轉變成新的結構，用已有的材料實現新的目標。例如，哺乳動物體內調控母親撫育后代的行為的荷爾蒙是催產素（oxytocin），一種古老的荷爾蒙，甚至在軟體動物體內也有發現，這種荷爾蒙被選擇承擔這項任務的歷史長達兩億年（Allman 2000：97，199）。我們現在知道，催產素也在配偶結合（pair bonding）的行為中起到關鍵作用，這提示我們，為了激發更廣泛的、超出母親—后代關系的助他行為，自然選擇在過去數百萬年中調整了它的作用（參見Gimpl and Fahrenholz 2001；Uvnäs-Moberg 2003）。如果親緣選擇為我們的遠古先祖提供了調控那些幫助親屬的行為所需要的心理和生理結構，那么在自然選擇的壓力驅使下，這些結構可以用于新的任務（最明顯的是，幫助非家庭成員的行為）。自然選擇的壓力可以有好幾種方式向這個方向去推動。

## 1.3　助他行為的演化過程：互利共生

有時候，某些目標對于生物個體是有利的，但這些生物不能獨自實現。一只獅子也許想用大象肉做晚餐——否則就得挨餓——但它單憑自己無法獵殺一頭大象。如果一群獅子都處于這種情形，它們可以通力合作，撲殺這頭大象。如果它們不合作，所有獅子都會挨餓，甚至面臨死亡。即使要撲殺的不是大象，而是可以獨力捕獲的動物，獅子們一起行動還是可以增加成功捕獵的概率，并且降低風險。顯而易見，自然選擇對引發這類合作行為的性狀青睞有加，這并不需要這些獅子之間具有血緣關系。這類助他行為有時被稱作互利共生（mutalism），有時則被稱作合作（cooperation）。然而，在日常語境中，“合作”也可以指許多其他類型的互利互惠行為（例如將會討論到的互惠和回報行為），因此，只在狹義上使用“合作”可能會引起混淆。而在學術圈外，“互利共生”并不常見，因此使用這個術語更恰當。[[19]](#_19_114)

基于互利共生的助他行為不是犧牲繁殖適應性的行為。一只不跟同伴合作的獅子，很可能把整個捕獵給搞砸了，由此降低了它自己的繁殖適應性。有的時候，不必所有的獅子都參與進來，捕獵也能成功，這時就會有一種選擇壓力促使獅子讓同伴繼續捕獵，而自己坐享其成。但是一般而言，參與捕獵的獅子越多，捕獲獵物的概率就越大（參與捕獵的獅子受傷的機會就越小），所以在大多數情況下，參與捕獵會更好地提高自身的繁殖適應性。

我們需要注意互利共生的一個特點（為了將它與助他行為進行對比）：互利共生的雙方并不需要有持續的關系。例如，為了趕走一只大型的危險動物，一群小鳥“群起而攻之”（互利行為的又一個例子），它們此時此地促進了彼此的適應性，即便事后鳥群就一哄而散，不再往來。因此，互利共生不屬于一種互惠關系。這種簡單的區別，我們可以用大衛·休謨的一個例子來說明。假設有兩個劃槳者同心協力把船劃到一個彼此都想去的地方。雙方沒有交換承諾，因為沒有必要：任何一方停止了劃槳，船就會原地打轉；在這種情況下，想到達目的地，就需要“他們都做好各自的工作，只有一個人劃槳的話，對雙方都沒有好處”［（1751）1983：95］。我們可以設想另一種情況作為對比：一個劃槳手就可以駕船。甲答應把乙載到乙的目的地，但前提是晚些時候乙也會把甲帶到甲想去的地方。這種安排就跟上例不同，包含著契約（雖然是一種隱含的契約）。前者屬于互利共生，而后者則屬于互惠。現在我們開始直接討論互惠。

## 1.4　助他行為的演化過程：直接的互惠

我們經常見到，一個人給另一個人提供幫助，幫助的價值遠遠超出他所付出的代價。假設有一種猴子容易受到某些寄生蟲的侵擾，那么能清除這些寄生蟲，對猴子而言具有非常重大的意義（甚至事關生死），而清除這些寄生蟲只需要花費半個小時。親緣選擇可以解釋為什么有的猴子愿意花費整個下午的時間，幫助它的家庭成員清理毛發；但它很難解釋為什么在自然環境中有的猴子愿意幫助非親屬成員。為非親屬成員清理毛發，帶來的好處遠遠多于清理者所付出的代價，盡管如此，它多少還是需要付出一定的成本：半個小時的時間，這足以使它找到充足的食物，或尋找到交配的機會。因此，為了沒有親緣關系的同類而做出犧牲，可以為它帶來什么好處呢？一個明顯的答案是，如果那些它幫忙清理的、沒有親緣關系的同類，在它清理完后（或者晚些時候）會反過來給它清理，那么這就是對大家都有利的安排。假如所有的猴子都參與這種合作，那么對任何一只猴子而言，總的好處就大于所付出的成本。第一個清楚地觀察到這個過程的人是羅伯特·特里弗斯（Robert Trivers 1971），他稱之為互惠利他主義（reciprocal altruism）。

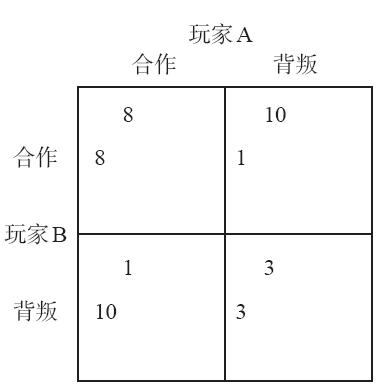
對于施惠者和受惠者都有利的幫助行為，特里弗斯主要的例子是珊瑚礁上的“清潔站”。在珊瑚礁上有一些小型的“清潔工小魚”（或小蝦），它們以獨特的游泳姿勢游向大魚，表示樂意為大魚清理身上的寄生蟲。如果這條大魚想要清理寄生蟲，就會張開口和魚鰓，讓這些清潔工小魚進去工作。這條大魚感覺差不多了，就會發出信號讓里面的小魚出來。這條宿主大魚隨時都可以在享受完小魚的清潔工作之后，再順勢吃掉它們，飽餐一頓。假如海底世界充滿了這種清潔小魚，那么，宿主大魚似乎就應該這么做。但是，珊瑚礁只能容納一定數量的清潔小魚，那么讓這種交易持續下去，就更有利于大魚的繁殖適應性。它能知道去哪里可以得到好的清潔服務，也知道這些小魚是非常穩定的清潔工，對它很有價值，勝過一餐飯。如果大魚為了長遠利益而放棄眼前的一餐飯，那么清潔小魚放棄了什么呢？小魚游向大魚（甚至進入到它的口中），冒著很大的風險；因此，它們放棄的是安全。如果小魚不承擔會被吃掉的風險，它們幾乎不可能得到什么好處（吃掉寄生蟲）。它們可以“耍詐”，趁大魚不注意在后者鰭上吃一小口（正如某些種類的“清潔魚模仿者”的做法一樣），由此以他人為代價來提高它們直接的總利益。然而這種做法是短視的。找到好的清潔工是不容易的，找到忠誠的顧客也是一樣。[[20]](#_20_112)

某些植物也有互惠利他主義式的成本—收益結構的關系。植物沒有能力做出欺詐行為，所以也沒有自然選擇的壓力來發展出覺察欺詐行為的能力。即使有些生物的認知能力具有可塑性，可以偶爾進行欺詐行為，互惠關系也不會很容易被利用。如果欺詐行為會破壞高度互利的關系，那么任何傾向欺詐的壓力很容易被反對欺詐的壓力克制。對處在持續交換關系中的雙方而言，要是情況一直如此，那么自然選擇完全不需誘使任何一方進行欺詐，或者教導它們如何應對欺詐。但因為只有當成本和收益在多種尺度上都達到平衡的時候，互惠交換行為才可以發展；也因為在現實世界中，價值很少是穩定的，所以，在環境因素發生變化時，互惠關系瓦解的可能性總是存在的。如果互惠關系的一方A表示，無論發生什么情況，它都愿意提供幫助，那么，回頭幫助A就可能不再符合B的利益。假若欺詐帶來的價值提高了（如果B可以吃掉A，而且B正面臨著突然的食物短缺），那么回頭幫助A就可能不再符合B的利益了。如果找到愿意提供幫助的新伙伴（雖然它們同樣在被騙之后就不再愿意），要付出的成本微不足道，那么再幫助A也同樣可能不符合B的利益。為了使得自然選擇青睞持續的交換關系，這些價值對雙方而言都必須是穩定的和對稱的。[[21]](#_21_113)有意思的是，在許多互惠關系中，欺詐對方（一旦它已經得到了好處）而且不受懲罰的可能性確實存在。因此，自然選擇偏向于在生物中發展出區分兩種欺詐的能力，一種是會導致長期損害的欺詐，另一種是有利可圖的欺詐。這又驅使生物發展出兩種能力，即對欺詐行為的敏感，以及對欺詐行為的反應能力。具有這些能力的生物之間的交換關系屬于經過算計的（calculated）互惠關系；處于這種關系中的個體，能夠感知到相互之間的成本—收益結構中發生的變化，并依此調整自己的回應方式（de Waal and Luttrell 1988）。

如果不回報行為要付出的代價非常大，不單單是喪失了持續的交互關系，那么互惠關系的成本—收益結構就可以被穩定下來。促使這種情況發生的可能原因之一是，個人在幫助了別人之后，如果沒有獲得回報就會懲罰對方。另一種可能原因則是，即使你本人沒有被利用，但你還是會懲罰（或拒絕幫助[[22]](#_22_113)）那些你觀察到有“不回報”特質的人。有人可能更進一步，會懲罰所有拒絕懲罰不幫助他人的人。這后一種懲罰性狀的發展可能會受到“高階背叛”（“higher-order defection”）的妨礙，因為相比自己回報別人也懲罰不回報者的人，那些同樣回報別人但不懲罰不回報者的人具有更高的適應性。羅伯特·博伊德與彼得·里查森（Robert Boyd and Peter Richerson 1992）證明了，只要群體的規模足夠小，小到所有群體成員都充分意識到不懲罰不回報之人所造成的消極后果，那么這就不是個難題。不過，他們論證道，為了解釋在一個更大群體中的懲罰性狀，我們必須訴諸文化的群組選擇（更多相關討論見1.7節）。

需要注意兩個重要因素。首先，這些“互惠利他主義者”并不是我所定義的“利他主義者”。畢竟，例子里的清潔小魚，并不具備做出日常意義上的利他或自私主義的行為所需要的心理條件；它們可能甚至都不具備進行行為的必要條件。其次，也許相對不那么明顯的是，這些助他的生物也并沒有表現出犧牲繁殖適應性的行為（因此，它們不是在“演化論意義上利他的”，參見Sober 1988）。在互惠交換中，雙方都沒有為了對方而損失其繁殖適應性。按照特里弗斯的定義，“利他主義行為”（他指的是助他行為）是那些“在表面上對該生物是有害的”（1971：35）的行為。但是顯然，表面上的犧牲繁殖適應性并不是真正的犧牲繁殖適應性，正如外表上像勞力士的手表并不是真正的勞力士手表。還有的人將“互惠利他主義”定義為短期的犧牲繁殖適應性行為。但是，為了長遠的價值而放棄短期價值并不是犧牲繁殖適應性，正如猴子為了找到樹上的果子而費力爬樹并不是犧牲繁殖適應性行為一樣。盡管經常有人主張，互惠利他主義和親緣選擇能夠共同解決所謂的演化利他主義難題，但如果（1）我們說的“利他主義”指的是犧牲繁殖適應性的行為（而不是表面的或短期的犧牲繁殖適應性行為），（2）“適應性”指的是總體的繁殖適應性，以及（3）“對演化利他主義這個難題的解決”指的是，解釋這樣的利他主義是如何可能的，那么我就不清楚還有什么理由表明這種常見主張是正確的。但是在另一方面，對助他行為的演化而言，互惠性很可能確實是一個重要的過程。出于這些原因，我更偏向使用互惠交換（reciprocal exchanges），或互惠性（reciprocity）這兩個術語，來指稱特里弗斯筆下的“互惠利他主義”。[[23]](#_23_113)

特里弗斯認為，我們可以用囚徒困境（博弈論研究者長久以來的至愛）來為自然界中的互惠交換構建模型。在囚徒困境（PD）中，兩個人要決定如何互動：他們可以同時采取合作，也可以同時背叛對方，或者可以一方合作而另一方背叛。但是他們需要同時做出決策，然后再比較決策結果。每一個可能的結果都與兩個參與者的“回報”相關（圖表1.1）。用習慣的標記法，8是R（reward for cooperation，給予合作的獎勵），10是T（temptation，誘惑），1是S（sucker’s reward，蠢貨的回報），3是P（punishment，兩者同時招供導致的懲罰）。囚徒困境要求T＞R＞P＞S，并且2R＞T+S。[[24]](#_24_111)



圖表1.1

如果你只能和另一個玩家一起玩一次這個游戲，要知道該怎么選擇并不容易。也許你覺得，相互合作會有好結果，可是一旦你選了“合作”，那你就可能被利用。你相信你的對手不會讓你只得到1？為了安全起見，也許最好還是選擇“背叛”，畢竟得到3起碼比得到1還是要好一些。當然，如果對方也是這么考慮的，最終你們會互相背叛。但是如果游戲是重復的，你和同一個對手一直玩，雖然不知道什么時候停止，那情況可就不同了。你需要制定一套策略，根據對手之前的選擇來做出決定。你可以決定先背叛幾次，然后再用合作來向她“道歉”。你也可以無視對方是怎么做的，一直選擇背叛。羅伯特·阿克塞爾羅德（Robert Axelrod 1984）使“以牙還牙”策略（Tit for Tat,TFT）聲名遠揚［雖然這最先是由安那托爾·拉波波特（Anatol Rapoport）提出的］。TFT極其簡單：第一輪與對方合作，從這往后一直重復對手上一輪的舉動。換言之，只要對方合作，你也跟著合作；永遠不要率先背叛對方，但如果對方選擇背叛，那你就立即跟著背叛。此外，如果雙方一直相互背叛，那就耐心地等到對方“道歉”為止（因為必須由她開始）。TFT是“友善的”，既不會被對方嚴重利用，也不利用對方。

現在回頭想想那只“搭便車”、不幫同伴清理毛發的猴子。假設它先請求一只沒有親緣關系的猴子（稱他為“A”）給她清理，A立刻幫助了她。隨后A請她幫忙清理，卻一無所獲。由于A根據“TFT”規則行事，A就不會再為她免費整理毛發，除非她決定以后也為A整理毛發。到目前為止，她的確是占了便宜，因為她得到一次無償的清理，而A卻一無所獲。但是如果我們考慮其A接下來的活動，她就不再是一個獲利者了。她從A那里得到了無償的毛發整理，我們再假設她也在族群里別的猴子那里都設法占到了同樣的便宜（每只猴子都根據TFT而行事）。從此之后，她的運氣就到頭了；沒有任何一只猴子愿意碰她（除了她的親屬。為了闡明我的觀點，這里假設單純依靠親屬的照料不足以抵御寄生蟲的滋擾）。同時，只要其他猴子還保持著相互交往，都樂于為彼此整理毛發。這只猴子最終將死于寄生蟲感染。搭便車就到此為止了！

人們對TFT有一個通常的誤解，認為TFT任何時候（或者在大多數情況下）都可以勝出。恰恰相反，TFT從來不會勝出。當對手選擇合作時，選擇背叛是在一輪博弈中拿到更多分數的唯一途徑——而根據TFT的定義，只有自己在上一輪的博弈中受到了這樣的對待，博弈者才會這樣做。TFT所能獲得最好結果是跟對手打平。然而，如果我們所說的“勝出”指的是別的意思，那TFT確實可以勝出。如果使用各種策略（或者各種策略的不同版本）相互爭斗，勝出者是整個競賽（假設這里不是“淘汰”賽）最后取得最高分的那位，那么TFT就可以帶來勝利。雖然與任何人作為對手，使用TFT最好的結果都是平局，但假如其他所有對手彼此博弈時都有輸有贏，那么采取TFT的人最終會成為贏家。在具有某些形式的競賽中，TFT通常可以取得極大的成功。

但是，實際情況比這里描述的要復雜得多。TFT的成功完全取決于游戲設定的方式，而且有很多理由認為，這個博弈游戲的規則未能模擬出真實世界里互惠交換的許多方面（進一步的討論，參見Hirshleifer and Martinez Coll 1988）。這里提供六個例子：

1.PD博弈的要點完全在于參與者要同時做出選擇，特里弗斯則強調，互惠交換關系的重要特征是時間上的延遲。要改進PD模型，可以引入交替的囚徒困境，其中博弈者做出決策之前已經知道對手的動向。

2.真實世界里的生物會出錯；在彼此之間的交流中容易發生誤解。對于兩個原本合作愉快的TFT博弈者來說，這是災難性的：如果一人認為對方背叛了自己，她就會立即背叛對方，導致彼此不斷選擇背叛。要改進模型，可以把“噪聲”（noise）引入博弈，從而有一定概率出現溝通失誤和意外發生。

3.從演化的角度看，有的策略比別的策略要付出更大的成本。采用TFT的博弈者需要有某種分辨的技能，而使用“堅持合作”（always cooperate，ALL C）或“堅持背叛”（always defect，ALL D）策略的人則不需要。因而，在只有TFT和ALL C兩類參與者的博弈競賽中（所有人都忙于跟別人合作），那些實行ALL C的人具有適應性優勢，最終獲得勝利。改進模型的相應方法是，對參與者的策略征收“復雜稅”（complexity tax）。

4.在跟別人往來之前，我們已經觀察到了他們的其他行為。換言之，在我們坐下來開始向對手采取行動之前，我們大概已經有一定的基礎來判斷她將采取什么策略。這很可能影響到我們對她所采取的策略，影響到我們在多大程度上可以容忍對方偶爾的“背叛”等。要改進模型，可以讓參與者樹立個人“聲望”，并根據對方的聲望調整自己的策略。［這種改進可以包括給博弈者提供“審查交易”（scrutiny deal），在這項交易中，一方可以付出一定的代價而得到潛在對手的各種信息。］

5.在標準的PD競賽中，一個參與者被迫與另一個人進行博弈，無論他多么反感自己使用的策略，但是，在真實生活中，我們常常可以主動叫停彼此的互動。與4的情況結合起來，我們還可以根據某人的聲望而選擇完全不同他打交道。要改進模型，可以把拒絕（再繼續）博弈作為博弈中的選項之一。

6.雖然有的人認為TFT會“懲罰”那些選擇背叛的對手，但其實這算不上真正的懲罰。它甚至算不上“以眼還眼”，因為要讓耍陰招的對手品嘗到一模一樣的苦果（即“蠢貨的懲罰”），唯一的方法就是迫使她合作，而你則選擇背叛。特里弗斯指出，人類的互惠交換關系具有“道德主義的侵略性”（“moralistic aggression”），這種侵略性不僅僅是TFT式的溫和反應，像“你要不改變自己的策略，我就會一直背叛你”，它表現為更積極的處罰（penalty）手段，即譴責、孤立乃至用暴力對待背叛者。為了改進之前的模型，可以允許參與者不僅能選擇背叛，還可以更進一步地懲罰他人（同時也要付出一定代價）。

在某些版本的PD競賽中，大部分上述修正都有所采用（雖然也許其中有的版本已經不能算是囚徒困境），博弈結果顯示TFT并不占據上風。首先，我們來考慮一下把“噪聲”引入交互環境的修正。正如我們注意到的，這對TFT的穩定發展意味著災難。另一個可選策略（它也曾風光一時）被稱為“PAVLOV”。“PAVLOV”［其能力由馬丁·諾瓦克與卡爾·西格蒙德（Martin Nowak and Karl Sigmund 1993）所發現］所遵循的策略是“獲勝就繼續，輸了就改變”，這里獲勝意味著得到R或T的回報，而失敗意味著得到S或P的回報。PAVLOV比TFT要更能承受意外。假設有兩個PAVLOV參與者，厄爾尼和伯特，正當他們都連續合作的時候，厄爾尼不小心按錯了背叛按鈕。伯特輸掉了這一輪，因此在下一輪游戲中就轉換到“背叛”策略。同時，由于厄爾尼贏了前一輪游戲，他就繼續使用背叛策略。由于兩個參與者都被對方背叛了，他們馬上就又回到了互相合作的策略。（一個驚嘆號顯示了噪聲的干擾。）

厄爾尼：……C C D!D C C……

伯特：……C C C D C C……

重回合作關系當然很好。但是，當PAVLOV意外發現了某個一視同仁的合作者時，讓我們看看會發生什么。在第一對參與者中，雜音干擾到PAVLOV；在第二對參與者中，雜音干擾到ALL C。

PAVLOV：...C C D!D D D...

ALL C：...C C C C C C...

PAVLOV：...C C C D D D...

ALL C：...C C D!C C C...

采用PAVLOV的博弈者會殘酷地剝削對方，直到有噪聲再次干擾這個過程。有些評論者隱約意識到，TFT不是PD游戲的終極策略，但他們還是贊同以下晦澀的觀點：使用某種“類似TFT”策略的人最后總是贏家。但是，假設我們認為TFT的“非剝削性”是其重要特征，那么PAVLOV很明顯就不能稱得上“類似TFT”。PAVLOV對那些愚蠢的善人毫不留情，這種特質對其成功貢獻巨大。[[25]](#_25_109)

雖然生物可能追求類似“獲勝就繼續，輸了就改變”的策略，以此應對周圍的環境（例如在做出覓食決策時），但指望這套策略發展成熟，從而支配智能生物的交往模式，則是錯誤的。為什么呢？采取PAVLOV所獲得優勝地位的一個重要原因在于，它能利用噪聲因素排除ALL C博弈者，并能夠把他們剝削至死。然而，發現沒有辨別能力的參與者并剝削它們，即便這樣做確實有生存優勢，那么生物的自然選擇也不大可能挑選出這種效率低下的機制：“耐心等待，直到你偶然使用了背叛的策略，然后看接下來會發生什么。”如果要先發現對別人都一視同仁的博弈者，才能獲得這種優勢，那么自然選擇大概更青睞更為直接的清除方式。有的生物可能會先故意背叛對方，然后觀望結果。但這種實驗性的背叛可能面臨嚴重的處罰。（如果你想知道某國法律是否處死叛國者，那么親自到那里去，犯下叛國罪然后觀望后果，顯然是非常糟糕的滿足好奇心的方式。）如果識別蠢貨和叛徒是重要的，那么最好的方法大概是觀察其他博弈者是如何相互交往的。顯然，收集信息需要付出一定的代價（從適應性的角度來看），但是事先知道對手將會采用什么樣的策略，這帶來的回報是相當可觀的。現在已經有一些研究者，嘗試模擬PD博弈的回應方式中的這一因素（例如，Sugden 1986；Pollock and Dugatkin 1992；Nowak and Sigmund 1998；Panchanathan and Boyd 2003，2004）。然而，這些研究通常只是用聲望的概念來反映一個人過去是否自發選擇過背叛。可是PAVLOV的成功表明，聲望也應該反映一個人是否是一個無條件合作者。

## 1.5　助他的演化過程：間接互惠

莎士比亞告訴我們，“凡人一生能享有的最純潔的財富就是毫無瑕疵的聲望”。達爾文以不那么夸張的語言說道，“……對贊美的熱愛和對榮譽的強烈情感，以及對輕蔑和惡名更強烈的恐懼”一起構成了“促進社會美德發展的強大動力”［（1879）2004：133，156］。把聲望引入我們的理解，我們就遠離了標準的互惠交換，而逐漸趨近于“間接互惠”。這是亞歷山大（Alexander：1987）的道德系統演化理論的核心，我也同意這個概念確實非常重要。在間接互惠的關系中，一個生物可以因為幫助其他生物而獲利，它得到的好處可能遠遠大于它之前的幫助所付出的成本，而且給它好處的并不一定是它的幫助對象。通過下述的例子，我們可以看到聲望與間接互惠密切相關：假設A對幾個同類成員非常慷慨，而C觀察到或聽說了A的慷慨行為。同時，C也聽說了B聲名狼藉。基于這些觀察所得的信息（也就是基于A和B各自的聲望），C選擇了A，而不是B，作為保持互惠關系的同伴。A付出代價幫助他人的行為獲得了實際的收益，但這些收益并不來自他幫助的對象。亞歷山大列舉了間接互惠的三種主要形式。

（1）對他人施加恩惠的個體，可能隨后被納入到對他有利的互惠關系中，因為對方已經觀察到他在直接互惠關系中的行為，并由此判斷他屬于可為自己帶來好處的交往者（施惠者的“聲望”或“地位”被加強了，這最終為他帶來好處）；（2）施惠的個體可能從群體中的所有或部分成員那里得到直接的補償（諸如金錢，獎章或作為英雄的社會地位），這些隨后提高了他（以及他的親屬）得到額外利益的可能性；（3）施惠的個體所得的獎勵，可能就在于使得他所幫助的群體的成功，反過來促進了他后代或近親的成功（1987：94）。

關于間接互惠的一個例子是阿拉伯鶇鹛（Arabian Babblers），阿莫茲·扎哈維（Amotz Zahavi：1997）曾經對其進行過多年的研究。鶇鹛是一種社會性鳥類，經常互相幫助：為其他鶇鹛喂食，放哨，等等。令扎哈維感到疑惑的不是這種互助行為本身，而是有些鶇鹛非常積極地提供幫助：例如，爭著搶著要放哨，或把食物強塞給那些不愿接受的鶇鹛。扎哈維提出的“不利條件原理”（Handicap Principle）認為，這些鶇鹛在努力提高它們在鳥群中的聲譽，用行為表示：“快看看我；我多強壯和自信，足以承受這樣巨大的犧牲！”如此展示強壯的身體，就有可能吸引潛在的配偶，震懾競爭者。所以這樣的行為盡管在表面上不利于鶇鹛，其實完全是促進繁殖適應性的。

再想想雄孔雀巨大而笨重的尾巴。乍看起來，它的存在威脅了自然選擇理論——查爾斯·達爾文曾經承認看到孔雀尾巴的羽毛就令他感到“惡心”（F.Darwin 1887：296）。然而達爾文大體上也解釋了這個現象。他意識到，雄孔雀尾巴進化的最主要自然選擇動力，是雌孔雀的配偶選擇。[[26]](#_26_109)如果雌性孔雀更喜歡配偶擁有巨大的扇形尾巴，那么雄孔雀最后將會有巨大的扇形尾巴；如果雌孔雀更喜歡配偶的尾巴是三角羽冠狀的，或是螺旋形的，紅的，白的，或者藍的，那么（其他條件相同的情況下）雄孔雀最終都會長出這些樣子的尾巴。在性選擇的過程中，配偶的偏好，或者對手之間的競爭，都可以產生新的性狀，而這些性狀原本對其所有者來說是有害的。[[27]](#_27_105)我這里并不是把一般的性選擇歸入到互惠范疇，只是舉例說明有時自然選擇會青睞代價不菲的助他性。雄性幫助雌性（比如送食物），后者因此回報以更大的利益，即繁殖后代的機會，這個例子屬于直接互惠。雄性對它的同伴都提供幫助，而雌性觀察到這些，并因此給予雄性更大的利益，即繁殖后代的機會（因此他們的兒子經常幫助同伴，而女兒則偏愛有助他行為的雄性），這個例子則屬于間接互惠。[[28]](#_28_105)正如性選擇可以產生極其笨重的身體性狀，例如雄孔雀的尾巴，它同樣可以產生代價很高的助他行為。如果好的聲望帶來的收益足夠多，那么自然也可以選擇發展能獲得好聲望的行為模式。如果好的聲望意味著與對群體成員一視同仁分享食物，那么一視同仁分享食物的性狀將得到發展；如果好的聲望意味著頭上戴個南瓜，頭戴南瓜這個性狀也會得到發展。此外，類似的情況也適用于懲罰，畢竟得到懲罰只是獎勵良好聲望的反面。如果某種促進自我利益的行為（或者任何類型的行為）受到嚴重懲罰，那么歸根結底，這就不再是促進自我利益的行為了（參見Boyd and Richerson 1992）。

一旦認識到間接互惠包含了由聲望和懲罰構成的系統，并且該系統所形成的壓力可以促進幾乎任何性狀的發展（包括代價極高的一視同仁的助他行為），那么我們也就能了解這個解釋框架的潛在重要性。為了再強調互惠關系的重要性，我們應該回顧一下亞里士多德在《政治篇》中的敏銳發現，即對于能夠以物易物的生物來說，事物都不只有一種功能：一支矛有利于打獵，但也可以用來交換其他物品；一個水罐可以很方便地裝水，但也可以用來交易；收集食物的技能可以滿足獲取營養的需要，但也可以用來交換其他的利益，不逐一列舉。顯而易見，增多資源的功能帶來的好處相當可觀。

## 1.6　助他的演化過程：群體選擇

對于直接與間接互惠交換而言，識別能力都至關重要。在直接互惠中，一方只幫助隨后會回幫他的另一方；在間接互惠中，一方會根據另一方過去的表現，來決定是援助還是懲罰。但是，在某些幫助行為的模式中，施惠者不需要這樣的識別能力。他們并不需要成為互惠型的施惠者；他們會幫助群體中的任何人，無論得到什么回報。乍看上去，自然選擇幾乎不可能青睞這樣的行為。我們對于這種行為的疑慮可以表達如下：這些施惠者看起來像真正的繁殖適應性犧牲者（f itness sacrif icers）。但是自然選擇怎么可能更偏愛犧牲繁殖適應性的生物，而不是增強繁殖適應性的生物呢？

在對群體（或多層次群體）選擇的辯護中，埃利奧特·索伯與大衛·斯隆·威爾遜（Elliott Sober and David Sloan Wilson 1998）說明了群體選擇如何起作用的。首先，讓我們設想一個沒有助他行為的群落。假設該群落包括200個個體。這個群落的成員是人類、青蛙、獅子、植物或者電腦程序都不重要。重要的是它們以一定的比率進行繁殖。假定基線比率（base-line rate）為1.1。“基線比率”指的是一個個體在沒有任何干涉（沒有得到他人的幫助，也沒有做出任何犧牲）的情況下繁殖的比率。因此，如果這個群體沒有發生什么特別狀況（沒有共享和犧牲等現象），它每一代的數量就會以10%的比例增長。（為了讓情況變得更簡單，我們假設這個群體是無性繁殖的，并且當新的一代產生時，老一代就立即消亡。）現在讓我們在這個群體中也放入少量會幫助他人的個體。這些施惠者給非親緣關系的成員提供幫助，也由此降低了自身的繁殖能力。假設它們犧牲了自身9%的繁殖能力，來為10個同伴各增加0.4的繁殖能力（這些利益的分配是均衡也是一視同仁的，無論受益者是否自己也幫助他人）。如果我們放置10個這樣的助他者在群體中，就總共有100個成員的繁殖能力提高了0.4。既然分配是均衡的，那么100個成員就有5個助他者。這個群體的繁殖能力最終是：5個助他者的繁殖速率是1.4；5個成員是1.0；95個沒有助他行為的個體繁殖速率是1.5；其繁殖速率是基線比率1.1。這就導致第二代的結果如下：

群體數量=259（12個助他者，247個非助他者）

首先，我們要注意的是助他者的巨大影響力。總群體數量的凈增長不再是10%的基線比率，而是達到了30%。其次，我們也需要注意到，助他者在新一代的群體數量比例：從5%降到4.6%。再經過一輪之后，第三代的情況是：

群體數量=330（14個助他者，316個非助他者）

助他個體所占比例進一步下降到4.2%。如果我們繼續下去，我們就會看到這個比例將持續下降。如果存在某種限制群體數量增長的環境壓力（實際上肯定有），那么助他性最終會消失，盡管它給群體帶來許多利益。助他性不僅不能在種群中確立下來，它還極易被淘汰。如果我們做個類似的測試，讓群體一開始有199個助他者，只有1個非助他者，這樣一來，盡管群體會爆發式地增長，但非助他者的群體數量比例會逐步而穩定地上升，而助他性狀則注定消亡。

現在就讓我們來看看群體選擇。假設群體還是從200個個體開始，其中10個是助他者，相關數值跟上次一樣。但是這一次，分成各具有100個個體的群體，這兩個群體至少有一段時間是相互隔離的。群體A只有100個非助他者，因此群體數量以1.1的比率增長，在第二代中增長到110個成員。群體B有10個助他者，帶給其他成員的利益同之前一樣（即另10個同伴繁殖率增長0.4，且均衡而一視同仁地進行分配利益）。這里有趣的地方是，相對于群體B的人口規模，助他者的比例從10%下降到9.4%，可是相對于整個群體（A+B），它們的比例從5%上升到5.4%（這是辛普森悖論的一個例子；參見Simpson 1951）。再過一輪，第三代的情況如下：

群體A　群體B

群體數量=121　群體數量=218

（所有都是非助他者）（19個助他者，199個非助他者）

相對B的總群體數量，助他者的比例下降的程度更大（只剩8.7%），但相對總群體，比例卻有所增長（達到5.6%）。如果一直這樣繁衍下去，結果就沒有什么新意；如果給群體數量增長設定一個限制，同上一種情況類似，助他性狀會逐漸消失。但是，假設在此之前，整個群體數量發生了特殊的變動。想象一下，整個群體數量（群體A和B加起來現在有339個成員）被混合在一起，并按比例削減到原來200個的規模，然后再分成兩個群體。由于助他者的比例有所增長，它們在新的起始群體中所占比例就大于剛開始時的比例。然后，我們進行一項重要的改變：允許成員選擇與它們更偏好的成員走到一起（沒有人可以強迫對方和自己加入同一群體）。這類偏好并不復雜，也許只不過是選擇可以受益更多的成員而已。對于任何成員（包括助他者）而言，加入由助他者構成的群體都可以獲得最大的好處。結果就是當兩個新群體形成時，助他者會有“抱團”的傾向。[[29]](#_29_103)助他者的比例在第三代中達到5.6%，導致200個成員中有11個助他者。讓我們假設所有助他者都聚集在同一群體中。

現在我們重新從兩個新群體開始：A和B。其他所有的情況都跟上次一樣，只不過這次群體B開始的時候有11個助他者，而不是10個。在再次改變分配之前，我們讓這兩個群體繁衍到第三代，相對總群體數量，助他者的比例就會超過6%。如果我們再把它們削減為兩個數量為100的群體，而且助他者還是會抱團，那么結果群體B的初始群體數量中就有12個助他者。如果我們讓群體這樣繁殖到第三代，那么助他者在總群體數量中的比例會超過7%，所以當再次削減群體數量，然后重新開始時，助他者的數量依然是會超出以往。剛開始的時候，這種變化是緩慢的，但是如果我們讓這種發展模式繼續下去，就可以得到一個曲線圖，顯示助他者數量越來越多。我這里的設定有意讓助他者沒那么容易成功。但要是它們能夠以再犧牲0.1的繁殖能力的代價來多幫助10個個體，讓受惠者都得到0.4的好處，那么助他者的數量增長會更快。與此相反，在前文所述的單群體模型（single-group model）中（所有200個成員都混在一組），這種更大的犧牲會加速助他者的消亡。

前文提到，在單個群體模型中，對其他成員一視同仁的助他者會極其容易被非助他者淘汰。那么，當非助他者出現在眾多助他者之中時，多群體模型（multi-group model）會發生什么變化？我們設計一個與前面類似的模型，但是群體A全是助他者，而群體B開始有10個非助他者（別忘了，單個群體模型只需要一個非助他者就能逐漸淘汰助他者）。于是情況如下：

第二代：

群體A　群體B

群體數量=500　群體數量=461

（全部是助他者）　（414個助他者，47個非助他者）

第三代：

群體A　群體B

群體數量=2500　群體數量=2125

（全部是助他者）　（1904個助他者，221個非助他者）

正如上文所述，助他行為可以極大地促進增長。乍看上去，非助他性狀看起來運作得很好：數量從10劇增到47最后到221。實際上，非助他者會慢慢占領群體B。但我們需要注意到，在全部人口中非助他者的比例發生的變化：從5%下降到4.9%再到4.8%。假如我們同先前一樣，削減總數量，組成兩個各有100個成員的群體（同時保留助他者和非助他者的比例），并且允許個體選擇跟誰走到一起。在這200個成員中，非助他者的數量會比剛開始時有所減少。跟上次一樣，所有個體（無論是非助他者和助他者）都想與助他者分到一組，因此非助他者就被趕到一起。如果這兩個新群體繁衍到第三代，非助他者的人口比例將會進一步降低，在下一輪重組時，它們后代數量就變得更少。如此往復，非助他性狀最終會消亡。

有人可能會抱怨說，上面這些步驟都只是在擺弄數字，以便得到想要的結果。這個看法倒也不是完全沒有道理。但是我們要記住，現在的目標是表明助他行為的確有可能通過生物自然選擇而發展形成。自然選擇可以用數以百萬年時間，產生一系列的生物個體，有幫助過度的，有幫助不足的，乃至使用錯誤的方式進行幫助的，這些助他性狀都以消亡告終。但如果夾雜在諸多不成功的性狀之間，有一個能取得恰當的平衡，并使助他行為得以發展，那么可以預計，這種性狀總有一天會出現的。大自然的耐性是很充足的。

同樣，讓生物繁衍三代然后改變其群體數量構成，這種設置也不應該只從字面上理解。等待三代的時間，只不過為了方便地說明在不同的群體中性狀的比例是如何上升或下降的。我的意思不是說，在現實世界里，生物群體都遵循這種“重組之后繁衍三代”的模式。重點在于助他者相互聯合的趨勢，這種趨勢直接蘊含在“人人都愛熱心人”這句格言中。包含有助他者的群體，比壓根沒有或只有很少助他者的群體更興旺發達。因此，助他性狀（即一視同仁、不求回報的助他行為）能夠得以發展。為了方便理解，我們可以設想，多群體模式的情況包括住在相鄰山谷的部落，或是聚居在不同谷倉的老鼠，它們周期性地聚在一起進行交配，并組成新的老鼠群體，但也僅此而已，這些都只是些簡單易懂的例子。同樣的模式也適用于看上去規模很大的單個群體，只要我們允許在這個群體里助他成員也有扎堆的傾向。

這里所描述的助他者果真犧牲了它們的適應性嗎？倘若如此，那么它們就具有真正的“利他的基因”了（這樣稱呼是為了和理查德·道金斯的著名比喻形成對比）。我們已經看到，在直接互惠助他的情形中，犧牲繁殖適應性只是表面的。珊瑚礁上的大魚犧牲了一頓免費午餐，這只是因為它得到了更有價值的長遠回報（免受寄生蟲滋擾之苦）。至于助他者，看起來確實是放棄了一定的繁殖適應性：以自身的繁殖適應性作為代價，它們促進了別人繁殖后代的機會。在最初的群體B（其中90個成員是非助他者，其適應性是1.1，而10個助他者具有1.0的適應性），誰具有更高的起始適應性？是非助他者。現在再算上那10個助他者提供的利益，分配給群體中的每個成員。平均而言，助他者的適應性將會上升到1.4，而非助他者的適應性將會上升到1.5。做個非助他者仍然更有好處。

要使助他者比非助他者有更高的適應性，唯一的計算方法是把群體A納入考慮之列。這樣一來，在由兩個群體組成的整體里，助他者的平均適應性是1.4，但非助他者的平均適應性就只有1.28。有人反駁說，這種計算并沒有相關性。索伯爾和威爾遜稱之為“平均計算的謬誤”的一例。他們認為，只關注整體的計算得出的適應性程度，模糊了這個變化的動態過程，而動態過程才是這里的關鍵。如果忽略了這兩個群體以不同的速率繁衍，而只是把它們視作一個數量為200的整體，其中10個成員的適應性數值為1.4，而另外190個成員為1.28，那么，我們就忽略了這些數值產生的原因。之所以有這些數值，是因為前者那10個成員把助他的利益分配給后者中的90個成員（以及他們本身），而剩下的100個成員則沒有獲得這些利益。

看起來唯一明智的結論是：（1）在一定意義上，這些助他者確實犧牲了適應性，但是在另一種意義上，它們并沒有；（2）有論證認為前一種角度更具有解釋效力。這樣的說法多少有些模棱兩可，但我覺得討論到此為止，也并無大礙，因為對目前的研究計劃而言，真正的犧牲繁殖適應性性狀并不是必要的。重要的是助他行為確實已經被自然選擇，而多群體的進路為這種選擇如何產生又提供了一個模型。

上文揭示了多群體模型如何使未經篩選的助他行為得以發展。顯然，如果這個模型里的助他行為是有所篩選的，那么助他行為就更容易得到發展。也許對于某些形式的助他性來說，生物群落的動態過程不足以促進其發展，但如果讓助他行為略帶有選擇性（比如不情愿幫助那些非助他者），那么助他性狀就會逐步發展。換言之，有時候群體選擇和互惠可以共同促進一定的助他行為，雖然兩者都無法單獨產生這種結果。我們沒有必要把這兩者視為相互排斥的過程。

也許在一種更為直截了當的意義上，互惠和群體選擇是可以兼容的：互惠只是群體選擇的一種形式。對于本章前面幾節的組織方式，索伯爾和威爾遜大概會表示反對，因為他們會認為親緣選擇、互利共生和互惠等都應該被納入群體選擇的范疇。在繼續本章內容之前，我要對我的概念分類做一點簡要的辯護。

先考慮直接的互惠利他主義。索伯爾和威爾遜主張，自然選擇的基本單元是性狀群體（trait group），即成員數量為n的群體（n＞1），各成員“相互影響彼此在某個性狀上的適應性，但不影響該群體以外的生物的適應性”（1998：92ff.）。由此他們得出結論，互惠利他主義只是群體選擇的一種特殊形式，其特殊之處就在于這種群體只有兩個成員。但是金·斯特爾尼（Kim Sterelny 1996）合理地指出，上述定義中的群體（包括互惠交換的雙方），同經常被當作群體選擇典范的所謂“超級生物”（尤其包括社會性昆蟲的群落），是兩種完全不同的類型。后一種群體的代表，常常具有極強的凝聚力和整體性，成員有共同的命運，并且這種群體擁有的適應性無法在個體層面上同等有效地得到解釋（例如，新出生的蟻后殺死自己姐妹的傾向）。和普通的生物體一樣，這種群體完全可以被視為是具有相當客觀性的選擇對象。相對而言，對于前一種群體，認為自然選擇發生在群體的層面，其實并非必需，因為這種群體層面上的描述，同個體層面上的描述是等價的。對于這類群體范疇，斯特爾尼倡導一種多元進路（Dugatkin and Reeve 1994），根據這種進路，自然選擇作用的對象到底是個體還是性狀群體（即自然選擇的過程是個體選擇或是群體選擇的過程），其實并沒有實質的區別，只有啟發價值的區別，而至于如何從中選擇，則完全依賴于“我們解釋和預測的便利”（1996：572）。

我同意斯特爾尼的觀點，所以也樂意承認，按照某種對于群體構成的寬泛理解，互惠關系可以算作是被群體選擇的結果，但把它說成是個體選擇的結果也同樣可行。約翰·梅納德·史密斯（John Maynard Smith）曾經說過，在這個問題上的爭論不是“關于這個世界是什么樣子的……在很大程度上，只是語義上的爭論，無法用經驗觀察來解決”（1998：639）。但是有一種群體選擇過程顯然不在此列，即斯特爾尼所說的“超級生物選擇”（1996：577）。有的人可能認為，人類的合作能力（比如道德）是超級生物選擇的產物，而其他人則指出，人類的合作能力可以只用互惠來解釋。這兩種假說截然不同，但是如果我們出于方法論上的考慮，決定把互惠（以及親緣選擇）歸入“群體選擇”，導致這兩個假說無法區分，那么，我們就不得不承認，這就為達成理論上的統一性而在解釋的細節方面造成了難以接受的損失。

## 1.7　人類超社會性的演化

在上文中，我回顧了助他性狀在生物自然選擇的壓力下形成和發展的四種過程：親緣選擇，互利共生，互惠交換（直接的和間接的），以及群體選擇。生物學家李·杜蓋金（Lee Dugatkin 1999）曾經稱之為“四條通向合作的小徑”；雖然我們應該對其他的可能途徑持開放態度（例如Connor 1995；參見Sachs et al.2004），但幾乎可以肯定，在動物世界里，這四種方式是助他性狀演化的最為重要的過程。然而，這四種過程是否足以說明人類生活的重要特征之一，超社會性，答案就很不明確了。有人覺得與靈長類近親相比，人類社會的復雜性反而與社會昆蟲的合作生活有更多的共同點，這種看法不無道理。然而，膜翅目昆蟲極端的助他性狀是出于不同尋常的基因關聯，人類則不同，其超社會性部分的原因是我們能夠以累積的方式逐代傳遞大量適應性的文化信息，這種能力是獨一無二的。雖然我討論的幾種互惠性幾乎肯定在人類發展早期發揮了巨大作用，并在人類的心智活動中留下了痕跡，但是現有的模型（Boyd and Richerson 1988，1989，1992）表明，互惠只有在規模相對較小的群體（類似黑猩猩群落）中才能發揮作用。對于本書的假說，這倒不是問題，因為道德確實很有可能在人類祖先規模較小的群落中棲居時開始形成和發展。然而，既然本章的總體目的是概括導致助他行為的各種過程，所以任何未能解釋超社會性的討論都是不充分的。

即便我們把懲罰非互惠者（non-reciprocators）的可能性包括在內，也無法改變一個明顯的事實：互惠本身不能解釋大規模的助他行為。因為我們要是這樣做了，就會面臨如下問題：為什么在演化的競爭中，執行懲罰的生物不會輸給“隨和的施惠者”？后者相互施惠，但不樂意花費精力懲罰他人。索伯爾和威爾遜會訴諸群體選擇來解釋懲罰性狀的發展：一群懲罰者可能勝過一群隨和的施惠者。有人可能會疑惑，到底為什么索伯爾和威爾遜的模型要訴諸懲罰。難道群體選擇不可以直接產生互惠的助他行為嗎？答案是：確實可以，但群體選擇通過懲罰性狀產生助他行為，比起直接產生助他行為，可能性要大得多。這是因為，雖然執行懲罰一般需要執行者付出一定代價，但通常來說代價并不高（與群體由此獲得的利益成正比）。假設你經營小本生意，某一天，有個人走進來向你索取月收入的20%。這對你來說犧牲確實非常大。但是，如果拒絕交款的懲罰是死亡，那么你交出月收入的20%就是明智的。比較一下你自己的犧牲，和這個敲詐者為了真的威脅到你而付出的代價。或許他要做的，只不過是時不時慢慢開車經過你的房子以示恐嚇而已。

犧牲繁殖適應性的性狀能通過群體選擇得以形成和發展，總是因為取得了競爭的勝利：提高個體繁殖適應性的動力，與促進群體利益的犧牲繁殖適應性的動力，往往背道而馳。索伯爾和威爾遜的主要貢獻是證明了前一種動力不一定總能獲勝。但顯然，只有在提高繁殖適應性的動力較小時，犧牲繁殖適應性的動力才更有可能勝出：換言之，同樣是使群體獲益的性狀，那些使個體遭受較小損失的，比起使個體遭受巨大損失的性狀更容易得到發展。而且，如果以失去社會的尊重（社會尊重就像某種神奇物質，可以給也可以不給）作為懲罰（McAdams 1997），或者以被排擠出持續的有利交換作為懲罰，那么懲罰就無須任何成本。

雖然沒有人否認基因的群體選擇的確可能發生，但是在人類的遺傳譜系中，它起到多大的作用，答案就不那么清楚了。這里主要的癥結在于群體身份對交配選擇的影響程度。例如，兩個激烈競爭的部落，倘若有任何聯姻或相互移居的情況發生，都不利于群體選擇在基因層面上發揮作用。哪怕這兩個部落拼得你死我活，但是只要勝利的成果是擄走戰敗方的婦女，或接納戰敗方的幸存者，那么基因選擇仍然會受到阻礙。

但是群體選擇并不只在基因層面上發生。要知道，自然選擇本質上來說與基因沒有關系。達爾文很好地闡述了他的理論，但他對現代遺傳學一無所知。只要存在性狀變異、可遺傳性以及決定于性狀的差異化繁殖（differential reproduction），那么就存在選擇（參見Lewontin 1970）。（如果選擇是被有目的的設計所引導，那么這就是人工選擇；否則就是自然選擇。）他的理論本身并沒有說性狀一定要編碼在基因里，或者能夠繁殖的東西必須是生物個體。后天習得的文化實踐可能導致不同群體的性狀差異，也可能在群體之間傳播，也可能影響到群體的延續和繁榮；因此，某些環境雖然沒有引發基因的群體選擇，但文化的群體選擇有可能在這樣的環境下發揮作用。[[30]](#_30_101)

要產生群體選擇，群體內部必須有一定的統一性，而且群體之間還必須有一定的差異性。雖然在基因的層面上，這些條件并非完全不可能滿足，但是群體間普遍的移居和通婚，使得在現實中滿足這些條件非常困難。但是在文化的層面上，這些標準似乎更容易滿足。人類學家喬·亨里奇和羅伯特·博伊德（Joe Henrich and Robert Boyd 1998）證明，跟群體的大多數人保持一致行為的傾向，有利于適應多變的環境，因為這種傾向可以讓人可靠而高效地接觸到那些更容易在周圍環境中取得成功的行為（同時參見Boyd and Richerson 1985）。模仿成功者，或者（結果經常是同一回事）模仿大多數人，可以使得個體“省掉獨自學習和試驗的代價，直接迅速掌握適應性行為”（Henrich and Boyd 2001：80）。因此，亨里奇和博伊德假定，基因演化在人類中形成了特定的心理機制：支持從眾傾向，而且這種特性是人類獨有的累積性文化的關鍵。尤其是再結合那些同運用懲罰策略有關的性狀，從眾傾向還可以解釋在群體之間的差異增大時，群體內部的差異是如何被壓制的。正如前文所述，懲罰在原則上可以產生群體中的任何行為，甚至是怪異或看似不適應的行為。但是，在這方面，文化性的群體選擇可能有重要作用：一旦存在由不同文化群體構成的集合種群，就會有一種選擇壓力，有利于那些“親社會”（prosocial）的文化性狀的延續和繁榮。從整體和長遠來看，一個以頭戴南瓜為文化價值系統核心的群體，會輸給一個看重內部和平，珍視為同伴幸福付出的自我犧牲的群體。所以，這是解釋助他性的另一套理論。

關于文化的群體選擇，還有一個重要的補充，即我們祖先在創造文化的過程中，極大地影響了他們生活的環境。文化性的群體選擇發生時，基因的個體選擇肯定也在進行，這樣一來，前者的結果將會極大地影響后者的進程。例如，像畜牧這樣的文化性生產活動可能促進乳糖可耐性的產生，從而影響到相應群體的基因構成。在西非，甘薯的耕作導致熱帶雨林被大量砍伐，產生了更多的死水，使蚊子得以大量繁殖，從而加劇罹患瘧疾的風險，進一步加大了鐮狀紅細胞等位基因的優勢，因為具有該基因的雜合子對瘧疾有抵抗性（Durham 1991）。還有人假設，古代文化創造的烹飪技術意味著消化食物所使用的能量更少，從而使得耗能巨大的人類大腦可以獲得爆炸性的發展（Aiello and Wheeler 1995）。如果這個假設是正確的，那么不單只是我們龐大的腦袋使得文化的發展成為可能，而且文化反過來也使得龐大的腦袋成為可能。

同樣，如果我們承認文化性的群體選擇可以產生某種社會風氣，其中非互惠的行為要受到懲罰，不愿意懲罰這種行為的人也可能被懲罰，不愿意從眾的人也可能要受到懲罰，那么，我們就必須要承認，該環境中的個體要面臨新的選擇壓力（這種壓力在文化演化之前并不存在）。所以基因層面的個體選擇可以產生某些心理性狀，使得個體在高度獎勵親社會行為的環境里更容易取得成功（參見Henrich and Boyd 2001）。這樣一來，個體的基因演化與文化性的群體演化就可以持續進行積極的相互作用，由此產生的生物不僅具有高度社會性，而且在很大程度上，其超社會性編碼在基因里。[[31]](#_31_97)由于人類是已知的唯一會發生文化演化的生物，這種文化—基因共同演化的過程，在對動物的助他行為的各種解釋中屬于特例。

## 1.8　結論與前瞻

除去所有這些促進助他行為發展的演化過程之外，我們還可以討論得更多，但是我這里所關心的并不是更為詳盡的概念分類。我現在轉向的問題是自然選擇形成助他性的手段。假設一群古代蜜蜂遇到某種壓力，這種壓力會促進額外的助他行為，而且親緣選擇可以用來解釋這種壓力。但了解這種壓力的來源并不能讓我們知道這種額外的助他性是如何形成的。我們知道，自然選擇不能用魔法創造助他性更強的蜜蜂；它只能作用于控制生物行為的（無論何種）既有機制，并加以調整或轉化，從而引發新的或更強的助他行為。出于這個原因，自然母親采用什么手段形成助他行為，并沒有一般性的答案，就像生物怎樣獲得移動能力也沒有一般性的答案。決定蜜蜂助他行為的機制與促成黑猩猩助他行為的機制也不大可能有相似之處。解釋這兩種助他行為的演化過程大體上是一樣的（例如，兩者都來源于親緣選擇），但是演化過程通過什么手段才得到各自的結果，則有巨大的差異。

本書接下來的三章將考察以下論題：在自然選擇為了人類相互幫助而青睞的諸多手段之中，有一種是“道德感”，意為一種做出道德判斷的能力。這種機制可能來自以上所述的某一種進程，也可能是幾種進程共同起作用的結果。我本人認為，如果確實有這樣的先天能力，那么它的出現最有可能是源于間接互惠。但是，堅持這個假說并不是本書的目的。（不過，我會在本書的4.6節討論這個問題。同時參見Joyce,forthcoming c。）我們還有一個需要優先處理的緊迫任務，要求我們暫時放下演化論的問題。倘若對什么是道德判斷缺乏可靠的理解，那么要解決道德判斷的能力如何演化這個問題，任何嘗試都行之不遠。（忽視了這一點，就好像寫一本名為《德性的起源》的書卻對什么是德性缺乏任何實質討論。）這就是下一章的目的，但是出于將在后文揭示的理由，我們要從討論親緣選擇和愛開始。

# 第二章　道德的本質

## 2.1　更新世的愛

盡管對于我們祖先社會生活的諸多細節，我們仍然所知甚少，但有一點是可以確定的：早在幾百萬年前，每個人都在其嬰兒時期得到了照顧。人科動物兩足直立行走，大腦體積較大，這兩個特征加重了嬰兒的無助處境。（現代人有三分之二的神經在嬰兒出生之后生長，因為能通過人類盆腔的大腦體積有限。）因為我們的人類祖先是哺乳動物，嬰兒依靠哺育喂養。這個顯而易見的事實意味著，我們知道至少有一個親屬必須密切地參與到幫助新生兒的任務中，即母親。父親的參與達到什么程度則可能存有爭議。如果假設猩猩與我們的祖先（500—700萬年前）比較相似，那我們就必須推斷，是母親主要承擔了養育幼兒的負擔，而父親的作用相對而言并不顯著。如果我們進一步假設，現代的狩獵—采集社會體現了離我們更近（例如10萬年前）的祖先的生活組織模式，我們就必須推斷，在某一時期事情發生了變化，“雄性親本的高度投入”成為了最大化繁殖適應性的更好模式（參見Wright 1994）。這個改變可能同由樹棲生活向草原生活的轉變有關。

母親—幼兒聯系毋庸置疑的重要性，足以使我們得出幾乎確鑿的結論：親緣選擇在我們遺傳中是一種重要的動力。無數的思想家都意識到，對親屬的偏愛是人性的一部分。休謨寫道：

自然給予所有動物對于后代相類似的偏袒。無助的嬰兒剛一誕生，哪怕在所有旁人眼中，它都只是個卑劣和可憐的造物，但它的父母都會以最深的感情對待它。對于父母來說，任何其他事物，無論多么完美無瑕，都不能與自己的孩子媲美。這種從人性原初構造中產生的激情本身就把價值賦予最不起眼的事物。［（1742）1987：162—163］

早在數世紀之前，斯多葛學派就意識到另外一個關鍵點：其他形式的情感可以從父母與子女的密切聯系中產生。西塞羅寫道：

斯多葛學派認為，認識到雙親對他們孩子的愛是自然產生的，這是非常重要的。從這一點出發，我們可以追溯所有人類社會的發展……自然要是希望我們繁衍生息，就不大可能不關心我們是否愛自己的后代……因此，自然賦予我們對子女的愛的沖動，就像自然賦予我們對痛苦的厭惡一樣明顯……這也是人類自然的、相互的同情心的根源……因此，我們自然地適合形成團體、群落和國家。［（45 BC）2001：84—85］

達爾文主義者可以這樣表達這一點：當我們的祖先仍然使用四條腿行走的時候，調節母親—后代聯系的神經機制就已經在哺乳動物的大腦中出現，因此生物自然選擇就有了基本素材，來讓我們發展更深遠的幫助關系。親緣選擇理論預測，這些進一步的幫助關系將出現在家庭成員之間，而許多的研究表明，現代人類的大腦確實為廣泛地幫助家庭成員的傾向提供了支持。

例如，人類所有社會里都有贈送禮物的行為。如果親緣選擇和互惠是重要的解釋過程，那么我們可以預期，只有在估計對方會回報自己的時候，贈禮者才會把禮物送給非親屬，但送禮給親屬則不需要這樣的前提條件。馬歇爾·薩林斯（Marshall Sahlins 1965）對人類學資料的全面回顧揭示的正是這一點：施惠者和受惠者的關系程度與對回贈的期望之間存在著明確的關系。蘇珊·埃薩克—維塔萊與邁克爾·麥圭爾（Susan Essock-Vitale and Michael McGuire 1980）考察了大量的人類學田野研究（從20世紀50年代一直到70年代），目的是為了評估那些研究是否符合下述預測：人們會給予親屬比非親屬更多的非互惠幫助；人們會給予親屬（近親尤甚）比非親屬更多的幫助；友誼關系是互惠性的；豐厚的和/或長期的禮物贈送現象更多來自親屬；需要大量非互惠幫助的個體會被逐步拋棄，先是被非血緣成員，接著是被遠親，最后則是被近親。除了少數的例外，這些預測都得到了證實。馬丁·戴利和馬戈·威爾遜（Martin Daly and Margo Wilson 1988）有一項著名的研究，他們發現，在不同的文化圈中，有血緣關系的家族成員間的謀殺率與有姻親關系的家族成員間的謀殺率都有一定的差別。拿破侖·沙尼翁和保羅·布戈斯（Napoleon Chagnon and Paul Bugos 1979）的研究顯示，在雅諾瑪人（Yanomamö）的復雜沖突中，通過參與者之間的基因聯系，人們可以更為準確地預測誰會選擇哪一方。甚至還有研究表明（Segal and Hershberger 1999），同卵雙胞胎比異卵雙胞胎更傾向于相互合作（當他們進行囚徒悖論游戲時）。要繼續列舉證據，支持“人類天生就有強烈的偏袒親屬的傾向”這個假說，并不是什么困難的事情（進一步的討論參見Barrett et al.2002，第三章）。

如果幫助親屬的傾向提高了人類繁殖的適應性（幫助的程度同關聯的緊密程度成正比），那么，自然選擇的過程對我們的大腦做了什么才能取得這個結果？我認為，部分關鍵答案其實簡單明了而且令人不難贊同：愛。父母對子女的愛，兄弟姐妹之間的愛，叔舅姑姨們對外甥子侄的愛，諸如此類。我們可以沒完沒了地爭論愛是什么東西，但這里讓我們先采納一個自然的答案：愛是一種情感（也許同時也是別的東西）。

這里首先需要澄清一點：也許從親緣選擇中產生的愛，對象非常有限，但這并不意味著愛除了是真摯的、非工具性的和真正以他人為對象的，還要有其他屬性。為了講清楚這一點，我們可以把這種情感與性愛做比較。如果艾米是一個普通的異性戀者，她只會對男人才會產生某種特定的情感，對女人則不會。即便有此限制，我們也沒有理由認為，艾米對男友的愛其實是犯了自我指向的（self-oriented，即心理學意義上的自私）道德錯誤。雖然這樣可能有點啰唆，但還是讓我重申一次之前的觀點：可以從演化的角度解釋愛（比如父親對子女的愛），并不意味著父親的愛是由優化自身總體繁殖適應性的潛意識欲望推動的，也不意味著這種愛“歸根結底是自私的”。一般而言，人類群體中的父親并不比軟體動物群體中的父親擁有更多關于整體繁殖適應性的潛意識動機。不了解你產生愛意的某個原因（比如不了解情感的演化起源），這不等于誤解了你的情感的真正對象，更不等于你在某種意義上是在自欺。

有差別的愛不僅有可能符合心理利他主義的，而且還有很好的理由認為，至少從親緣選擇的角度看，這種心理機制可以讓自然選擇更好地起作用。非利他主義的愛（如果這樣說不自相矛盾的話）是指一個人對她愛的人的關心只是衍生性的，因為他人的福利可以反過來促進她自己的幸福，而自己的幸福才是她行為的最終動機（一個經濟學家也許會說，被愛者的福利是施愛者的效用函數的一個影響變量）。然而，親緣選擇最有趣的一點就在于它可以促使人們為了親屬的福利而損害或犧牲自己的利益，哪怕沒有任何回報。如果愛歸根結底的是基于自身的利益，這樣的動機根本無法支持犧牲行為。以基因的利益為終極目標的愛，倒是有可能支持犧牲行為——設想一下，人類被設計為能夠為自己的整體繁殖適應性而斟酌考慮，還幸運地有能力依據漢密爾頓規則（Hamilton’s Rule）進行計算和行動。但沒有人會真的這樣認為；提議基因這一概念是我們固有的，這顯然非常荒謬。那么，這件不可思議的好事是如何演化形成的呢？如果自然選擇想要人們為了基因的利益而行動，并使用愛這種情感作為達到該目標的近因機制，那么，以基因為導向的愛行不通，而自我導向的愛（盡管這種情感毋庸置疑非常重要）也不會產生最大化總體繁殖適應性所需要的自我犧牲行為。既然某些重要的基因存在于他人（尤其是親屬）體內，那么自然的解決方案就是創造一種（有差別的和有條件的）非衍生性的、以他人為導向的愛。（參見Kitcher 1998，2005）

所有的經驗證據都表明，人類的行為經常是出于對他人的真誠關心，而非歸結于自私自利（Pilliavin and Charng 1990；Batson 1991，2000；Ray 1998）。在這個問題上，有些證據可能出人意料：人類非常樂意懲罰違規者，寧可付出一定的物質代價——哪怕自己只是個旁觀者，而且也沒有受到違規行為的影響（Fehr and Fischbacher 2004；Knutson 2004；Carpenter et al.,2004）。在我看來（如果這里容許人身攻擊的話），與此對立的觀點一般來自對人類行為的某種悲觀的犬儒態度，而不是來自于任何真實的經驗根據。人類這種生物確實時常作奸犯科，以自我為中心，自我欺騙等（對此沒有人否認），但鼓吹以下普遍的心理學觀點，即所有的人類行為都由內心貪圖利益的小人所引導，這是錯誤的。沒有任何先驗的理由能支持這個主張，也不存在有利于該主張的經驗證據；事實的情況恰恰與此相反。

有了產生這種利他主義情感和動機的神經機制，生物的自然選擇就可以通過調整這些機制來實現新的目標。例如，可以利用愛親屬的能力來創造愛某些非親屬的能力。羅曼蒂克的愛情可能就是這種愛的結果，把夫妻聯結起來，還（這也許是一種更犬儒的觀點）激發了對通奸特殊的憤慨（Griff iths 1997：119）。曾經有一種流行的觀點認為，愛情的觀念在中世紀歐洲才出現；一位19世紀的評論家，亨利·芬克（Henry Finck 1887）甚至把愛情出現的日期追溯到1274年5月1日，那天年輕的但丁狂熱地愛上了比阿特麗斯。雖然在中世紀的歐洲，愛情在文化上得到了顯著的表達，但更寬泛意義上的愛情可能在數千年前就已經產生了，這與拉·羅什富科的玩笑相反——他認為人們要是沒讀過有關愛情的書就不會墜入愛河。根據一項目前為止對于愛情最詳盡的跨文化研究，89%的社會具有愛情觀念，而研究人員還認為，其余的社會里很有可能也有愛情，只是人們不怎么談論愛情，所以并未引起人類學家的注意（Jankowiak and Fisher 1992；同時參見Jankowiak 1995）。在自然選擇創造愛情的能力時，很有可能利用了已有的、產生親社會性的動機神經機制。（我已經討論過催產素的作用，參見1.2節）

我為什么要討論親緣選擇和愛情呢？因為我想要證明（我認為上文的討論已經足夠證明），對于人類為什么對某些他人懷有“親社會情感”（比如愛情）的問題，親緣選擇提供了一個既便利又簡單，還有經驗證據支持的演化解釋。但是我在這里真正想強調的是，這個答案還遠遠沒有牽涉對道德的解釋。問題不在于親緣選擇產生的幫助行為偏袒親屬，因為正如前文所述，親緣選擇也可以解釋針對非親屬的利他主義幫助。我們暫且放下親屬/非親屬的問題；關鍵并不在此。假設我們又通過訴諸互利共生和群體選擇，為一系列的人類親社會性情感（愛，同情，利他主義）提出一個可靠的演化解釋，這些親社會性情感可以廣泛地，甚至普遍地擴展到我們的同胞。不少理論家似乎想下結論說，這樣一來我們就發現了人類道德的起源，因此可以斷定道德是人性的一部分。我想要著重強調的是：這種觀點大錯特錯。

## 2.2　抑制和禁止

幾天前，我在一位朋友家里吃午餐。午餐很豐盛，將近結束時，他對我說：“別覺得你有義務把菜都吃完。”我邊吃邊答道：“我不覺得有義務要吃完；我是真想吃完。”這簡單的交流突出了一件明顯卻常被忽視的事情：你做某件事是因為你想做還是因為你認為應該去做，這兩者是非常不同的。[[32]](#_32_95)不難想象有一群愿望相同的人：他們都想過和平融洽的生活，不知暴力為何物。四處都是友好善良、相互扶持的人們。然而，沒有理由認為這里存在道德判斷。這些人們有對于殺人、偷盜等行為的抑制。他們做夢都不會夢到這些行為；他們也不會想去做。但是我們不必認為他們應該具有禁止的觀念：“人不應該殺人或偷盜，因為這些行為是錯誤的。”但是，做出道德判斷的條件之一就是理解禁止的能力。

這不能同伊曼努爾·康德的著名觀點混為一談。康德認為，被親社會性情感激發的行為在道德上不是可敬的［Kant（1783）2002：199—200］。我非常愿意接受常識的立場，反對康德的這個說法。別人的行為動機如果是愛、同情或利他主義，我們就愿意在道德上給予贊揚。我其實還認為，一個人的動機要是出于明確的道德計算，而不是直接的同情，這體現的其實是一種惡德。所以，這里并不是否認上述設想的人們值得我們稱贊，也不是否認他們在某種意義上是具有美德的。我的觀點爭議程度要小得多：一個人如果只是出于愛或利他主義的動機而行動，他并不由此就做出了道德判斷（假設這些情感本身并不必然包括這樣的判斷——這么假設看起來是沒有問題的）。如果我們研究的問題是道德判斷（從道德的角度思考彼此和世界）是否屬于人性的一部分，那么解釋自然選擇如何使人類形成對他人的利他的、同情的和愛的傾向之類，以及自然選擇如何產生好人（或者也可以說是有美德之人），其實是偏離了目標。主張人類“天生就具有道德感”，但提供的證據只是人類祖先以道德上令人（我們）稱道的方式行動，這就好像主張人類天生就具有簡單的算術能力，但證據只不過是人類祖先有十個手指。

雖然我著重強調了這層區別，但我并不是要說同情和其他的社會情感對我們的道德生活不重要；正如我們稍后將在合適的時候所看到的，情感毫無疑問對于人類的道德至關重要。此外，我也不是特別反對把“道德情感”這個標簽貼到愛、同情與利他主義等親社會性的情感上，或者把“美德”這個標簽貼到親社會的行為上，這些行為可能出于以上這些動機（雖然在哲學傳統中美德的含義遠不止于此）；但做出了這樣的讓步，就必須要聲明，道德判斷的能力對于這些道德情感或美德而言并不是必需的。當然了，這些道德情感時常伴隨著道德判斷：當你愛上一個人，你通常會做出傷害對方在道德上是錯誤的這一判斷。但是道德判斷并不是道德情感的必要成分，因為不難設想，有一個人對另一個人的感情符合愛的標準（而且完全不想傷害對方），但是她毫無避免傷害對方的義務感。情感的本質在下一章將會得到更詳細的討論。這里所強調的只是一種解釋，說明人類如何形成親社會的傾向和厭惡（無論這些傾向和厭惡基于愛和同情，還是憤怒和反感），并不等于解釋了人類如何發展出判斷道德對錯的能力，并且在這個意義上，也不同于解釋了人類的先天道德能力。這充其量只是解釋的開端。

## 2.3　何為道德判斷？

我一再重申，“只是想要”幫助他人（或“只是想要”看到那些不愿意幫助他人的人受傷害），這本身并不包括道德判斷。現在我們必須直接面對問題：道德判斷究竟是什么？一個人判斷某件事情是被禁止的，從而決定不做（而不是因為非道德性的抑制而不那樣做），這是怎么一回事？我們無須對道德判斷進行徹底的概念分析（我們不必在所有可能世界里都能準確找到道德判斷），因為對本書的計劃而言，了解了道德思維的行為后果就足夠了。但無論如何，為了明確目標，我們還是需要做些工作，否則我們會遭受無知之苦（正如在柏拉圖的《美諾篇》里，蘇格拉底說的），不了解討論的對象，因而無法進行任何有成效的探討：

美諾：蘇格拉底，你能否告訴我，美德是從教育中習得的，還是通過實踐而掌握的？如果美德既不能從教育或實踐中獲得，那么它是否來自于人的天性，或者別的什么地方？

蘇格拉底：……如果你覺得我能回答你的問題，那真是太抬舉我了。實際上，我根本不知道美德是什么，遑論美德是否可以教了。

蘇格拉底自稱無知時，不總是真誠的，對于目前的問題，我們的感覺可能也類似：我們不都知道道德判斷是什么嗎？但是我認為，這個問題極其復雜，而且根據我的經驗，任何談及這個問題的具體觀點都會招來批評。對于很多（真實的或想象的）例子，哲學家和普通人一樣都不確定有沒有做出道德判斷。（例如，某個精神病患者同意“偷竊是錯的”，卻絲毫不想去遵守。）哲學家們甚至在道德判斷是什么類型的東西這個問題上也無法達成共識。有些人堅稱，道德判斷屬于心靈事件，有些則認為，道德判斷是一種言辭。也許在這個問題上，最穩妥的方案是接受日常用法，容許道德判斷可以是兩者中的任意一種。但即便我們聚焦于作為言辭的道德判斷，這里的分歧依然很多。道德判斷是什么類型的語言表達？有些哲學家認為道德判斷表達的是命令；有些認為道德判斷是我們表露感情的方式。還有些人聲稱，道德判斷是對事實的報告。可究竟是哪一種事實呢？有的觀點認為，道德判斷報告的是有關說話人的事實，或有關說話人所屬文化的事實，或關于上帝命令的事實；有些人（道德實在論者）則認為，道德判斷報告的是獨立于心靈的客觀事實。但又是哪一種的獨立于心靈的事實呢？對于這個問題，也有很多不同的答案。元倫理學這個領域（它不可避免地要在這些相互爭鳴的立場中有所選擇）有時會讓人困惑不已。

我們顯然沒有必要在這里一舉解決所有爭議（這是一件好事，否則對那些問題的討論將會填滿這本同時也涉及其他主題的書）。但是，我希望排除一些理論選項（其中最突出的是“純粹的非認知主義”），還將用一定的篇幅來強調一些我認為是道德判斷的特征，這樣我們就不會把它們同其他形式的價值判斷相混淆。本章余下的內容就是處理這些任務。第三章討論的是道德情感，然后我將回到關于演化的問題。

當我們把道德判斷設想為某種公共的言論，引發元倫理學爭議的問題就浮現了：“道德判斷表達的是什么類型的心靈狀態？”我們必須謹慎處理“表達”一詞的意思。這里我們并不是在問是什么類型的心靈狀態引起道德言論。誠然，有時在別的問題上，我們確實在這個意義上使用“表達”。艾米莉踢翻了她弟弟做的沙子城堡，我們說這樣做表達了她的憤怒，意思是說憤怒引起了她的行為，或者說，憤怒是解釋她的行為的一個重要因素。如果艾米莉其實一點也不憤怒，我們就不得不否定這個解釋。然而，我們常用的“表達”一詞意思與此不同。當艾米莉稍后為踢翻沙堡向她弟弟道歉，她表達的是負罪感的情感。假設艾米莉的道歉是不真誠的，不帶有絲毫負罪感之情，但這并沒有改變她做出道歉的事實。一個不真誠的道歉仍然無損其成為一個道歉（正如一個不真誠的承諾仍然是一個承諾，一個不真誠的主張仍然不失為一個主張）。雖然艾米莉的道歉沒有誠意，她仍然表達了悔意，因為道歉就是悔意的表達。這里的“表達”并不是指艾米莉與她的心靈狀態之間的解釋性或因果性的聯系，而是指一種更加復雜的關系，這種關系包含了艾米莉、她的弟弟以及一系列的語言習慣，根據這些語言習慣，當一個人在恰當的環境下說出“我對不起”，那么她就已經（至少）表達了悔恨之情。[[33]](#_33_95)

純粹的非認知主義認為，道德判斷（作為言論）沒有表達信念，而是另外的一種言語行為。站在語義學立場上，非認知主義認為，評價性的謂詞（“……是錯誤的”，“……是善的”，等等）只有在語法意義上才算是謂詞；在更深的層面上（當我們認識到一個道德判斷的“真正意義”時），這些謂詞就消失了。非認知主義哲學家A.J.艾耶爾（A.J.Ayer）在20世紀30年代有一個非常著名的論斷，即“偷錢是錯誤的”，這個判斷并不是表達一個或真或假的命題，而是類似一個人用某種聲調（用符號表示就是感嘆號）說：“偷錢！！”以此來表示他的不贊成的感受［Ayer（1936）1971：110］。我認為，非認知主義有些地方是對的，但純粹的非認知主義則過于極端。例如我們可以比較一下，同樣是稱呼一個人，“德國佬”和“德國人”這兩個詞有何不同。說“漢斯是一個德國人”是把漢斯描述為具有特定國籍的人。說“漢斯是一個德國佬”，這不僅指出了漢斯的國籍，還表達了一種貶損的態度。說話者的輕蔑態度，不僅僅是聽者從過去的觀察而得出的預期，更是隱含在這個詞語當中的一種根深蒂固的語言習慣——對一個不知道其輕蔑意涵的人，我們可以說他根本沒有恰當地理解這個詞，哪怕他能把這個詞用于所有德國人且只用于德國人。宣布“漢斯是個德國佬”，這個言語行為顯然不能簡單等同于用嘲笑、奚落的語氣說“漢斯！！”，因為：（1）說話人同時表達了漢斯是德國人這個信念，（2）他對漢斯表達的輕蔑不僅僅針對漢斯，而且針對所有德國人。因此，對“……是一個德國佬”這個謂詞做艾耶爾式的解釋完全不可行。這也讓人們懷疑以那種方式解釋道德謂詞也是不可行的，理由是一樣的。或許當我們說“偷錢是錯誤的”時，我們既做出了一個關于偷錢的斷言，同時也表達了對譴責這個行為的準則的一種意動性接受（conative acceptance）。［在選擇表述意動性因素的方式上，我參照了Gibbard（1990）的觀點。］有一個簡單的論證說服我接受了這個觀點。

按照我們現在對于表達關系的理解，如果人們在環境C中說出一類句子S，該類句子的功能是表達某種心靈狀態M，那么顯然我們就可以預期，某人在具體環境C中說了S，接著馬上加上一句，“但我并不具有M”[[34]](#_34_93)，這就會令聽眾感到困惑。因此對于某些特定的S、C和M而言，如果確實造成了這種困惑，而且沒有其他明顯的方法能解惑，那么，我們就應該將這看作是S在C中表達M的證據。設想一下，在某種情境中，單單一個“對不起”就能夠構成有效的道歉行為，但某人在此說道：

（1）對不起。但我沒有絲毫悔意。

這肯定會引起聽眾的困惑；用J.L.奧斯丁的話來說，這是“一句說不通的話”（1971：18）。聽者并不會認為對方正確表達了道歉之意。（1）不僅聽起來“怪怪的”，因為說“對不起；但是我并不打算梳羊毛”（在跟羊似乎沒什么關系的情境里），也同樣很奇怪；“聽起來怪怪的”的句子到處都是，但（1）令人費解的緣故很特別。第一個句子表達了一種心靈狀態，而第二個句子則報告了一種心靈狀態——這里的古怪之處在于表達和報告的內容不一致。原本單獨說出第一句話能實現的言語行為，似乎被第二句話給消除了。G.E.摩爾讓我們意識到某些話語的古怪特性，例如“這只貓在墊子上；但是我并不相信”。這就是后來為人所知的摩爾悖論。之所以稱為悖論，是因為雖然這對句子并無矛盾（這只貓在墊子上而我對此不相信，這是完全可能的），但是同時說出二者會使得第一個句子的語言行為無效，也讓聽眾感到困惑，不知應該如何看待說話者所持的對這只貓是否在墊子上的態度。這就造成了“一類特殊的胡言亂語”［Austin（1970）1990：112］。我認為下面的句子也屬于此類現象。

（2）漢斯是一個德國佬。但是我并沒有鄙視漢斯或這個國家的其他人。

（3）謝謝你。但是我對你并沒有感激之情。

我們或許會將“但是我并沒有……”這部分理解為給前一句話制造了虛構的語境。要是聽到了其中的一句話，我們可能先遲疑一會，然后回應道：“你把漢斯叫作德國佬，只是在開玩笑吧？”（“你說謝謝我，只是在開玩笑吧？”，諸如此類）盡管這類觀察有些粗略，但我覺得是很有用的。如果“但是我沒有心靈狀態M”這個陳述足以削弱前面一句話的嚴肅性（也就是說，可以被顯而易見地解釋成是表明了前一句話的虛構或玩笑語境），那么，我們可以合理地假定前一句話必須起到表達M的功能。讓我們再用兩個道德判斷來檢驗一下這種假定。

（4）埃爾金大理石雕應該歸還希臘。但是我不認為它們應該歸還希臘。

（5）希特勒卑鄙邪惡。但是我不認為他卑鄙邪惡。

要是沒有后一句評論，這里的前一句話本會很容易被理解為道德判斷。我認為說出這兩組句子極有可能與說出（1），（2）或（3）引發同樣的反應。因此，（4）或（5）的前一句話，如果在被當作道德判斷的環境中自然而然地說出來（即不以諷刺的語氣說出來，也不在話劇表演中說出來，等等），就應該被看作是信念的表達——看作是肯定的斷言（assertion）。這就構成了反對純粹認知主義的證據。但是現在考慮下面的兩對句子：

（6）埃爾金大理石雕應該歸還希臘。但是我接受的道德標準都不要求把它們歸還給希臘。

（7）希特勒卑鄙邪惡。但是我接受的道德標準都不譴責他的品格或行為。

這些句子顯然也很古怪，說出來會遭到質疑，所以在某種意義上它們也是“說不通的”。我謹慎地傾向于認為，這些句子也是摩爾悖論的例子，因此，道德判斷是肯定的斷言但又不僅僅是肯定的斷言，它們同時表達了信念和意動性的狀態（比如接受）。

如果我們仍然對她是否做出了道德判斷這件事存有懷疑，那么我們也許可設想一種情況，在這種情況中A和B不相關。我們需要問問自己，做這樣的想象時，我們想到的是不是另類的少數群體，是不是人們在開玩笑，說話人是不是用的嘲弄或反諷的語氣，或者我們是否想通過實例來引入新的語言習慣，或者我們是否在設想一個跟現實世界有不同語言習慣的世界。如果我們的想象包含任何這類的特征，那這就不會對我的觀點構成反例（這里涉及問題比我們現在能探討的更為復雜。進一步的討論參見Copp 2001和Joyce,forthcoming a）。例如，很少有人會懷疑“slut”（蕩婦）一詞在英語里有輕蔑含義。但一個人可以通過突然改變語氣或眉毛上挑等，表明現在遵循的是開玩笑的習慣，跟親密的女性朋友鬧著玩說“哎喲，你真是個小蕩婦”，而且沒有絲毫冒犯的意味。然而，這種情況并沒有削弱以下主張：根據某種現有的語言習慣，“蕩婦”一詞表達了輕蔑的含義。即使當我們以開玩笑的方式使用“蕩婦”，它仍然是一個表達輕蔑的詞語，歸根結底，正是因為如此這個玩笑才有趣。

所以我認同的觀點是：（作為言語行為的）道德判斷既表達了信念，也表達了不同于信念的意動性狀態。因為堅持前半部分，我反對純粹的非認知主義，比如艾耶爾的觀點。（下一章將進一步討論我的理據。）因為堅持后半部分，我反對純粹的認知主義。純粹的認知主義認為說某人是個惡徒只不過把一類特征——類似個子高，或是挪威人——歸到這個人身上，并不以此表達對這個人的態度。我反對這種看法，因為我認為它明顯是不正確的。說瑪麗道德敗壞并不同于做出一個中立的陳述，像“她沒有付錢就從書店拿走了一本書”。說后一句話時，說話人沒有表明對瑪麗的態度；他完全有可能是一個無政府主義者，熱忱地支持入店行竊的行為。但是，如果是以嚴肅的口吻說出“瑪麗道德敗壞”這個句子（不是開玩笑，也不是使用“敗壞”一詞的其他含義，等等），那種中立解釋就是行不通的；我們知道說話者確實是在譴責瑪麗的行為。

要是我們把表達關系當成因果性的，這個結論似乎沒什么吸引力，因為這樣一來，這種“雙重判斷”全都要產生于某種兼具信念和欲望兩方面特征的心靈狀態——用J.E.J.奧爾瑟姆（J.E.J.Altham 1986）發明的術語就是信欲（besire），而反感這種心理學的人會否定我們的理論前提。但只要我們意識到相關的表達關系并不是因果性的，就可以明白這種結論是得不出來的。一句話可以同時表達信念和態度，這跟說話者的信念狀態與他的態度狀態之間存在何種模態關系是沒有關系的。一旦排除了這個顧慮，下面這種語言習慣也就不會遇到什么哲學上的困難：依據這個習慣，某人在環境C說出S同時表達了兩種心靈狀態。[[35]](#_35_91)

這個結論值得留意，在后面的論證中，對純粹非認知主義的否定將作為前提再次出現。但是，上面的說法都還沒有揭示道德判斷有什么特別之處。說希特勒“邪惡”，和稱一匹老馬是“a nag”（“一匹駑馬”），我們目前為止的分析都一樣適用。因此可以看到，現在得到結論甚為簡單。說一匹馬是“a nag”是一種評價；在英語的習慣用法里，這表達了對該動物略微厭惡的態度。但是，道德判斷當然有更多的實踐內涵，并不僅是根據某些語言約定，選擇某個相關的謂詞可以讓我們既表達態度也交流信念。道德判斷的實踐影響肯定不只是語言上的花樣。

## 2.4　道德判斷的實踐影響力

我們現在依然在處理“何為道德判斷？”這一問題。道德判斷有一個必須要強調和探討的重要特征，即它對于我們的慎思和交往的所謂實踐影響力的程度：稱一個行為是“道德上正確的”、“有德的”、“錯誤的”或“正義的”，（據說）就是在把我們的注意力引向某種無法合理忽略或回避的慎思考慮。

但是，單單只是“表達（對某物）的感覺”，或表明自己的態度，這些根本不是提出慎思考慮——只要我們想想，感覺的表達（例如“太好吃了！”）總能夠被斷言性的報告代替（“我在享受這些食物”），這一點就很明顯了。表達自己的感覺，和報告自己的感覺，是兩類不同的語言活動（在過山車上，一個人尖叫并不等于他在很冷靜地陳述道：“我現在感覺非常興奮”），但兩者為聽眾提供了關于說話者內在狀態的幾乎一樣的信息。然而，單是報告感覺，這本身從來不能為聽眾提供實踐考慮。假設羅杰是一個反狩獵的狂熱積極分子。一群追獵狐貍的人跑過他面前，他對這些人說：“你們的活動我不贊同。”如果獵人略有困惑地回應道：“是的，那又怎么樣？”我們很難說獵人們做錯了。即便羅杰不是通過報告，而是通過吼叫表達他的不贊同：“你們這些殺害狐貍的獵人，呸！”，同樣的回應也還是在情理之中的。如果前面的感覺報告沒有給這些獵人提出（或意圖提供）停止獵殺的理由，那么后一種感覺表達也不能。要是羅杰認為后者能夠提供理由的話，那是非常奇怪的。如果道德判斷只不過是說話者意動性態度的表達，那么它們同樣與他人的慎思不相關（除非聽者恰巧關心說話者的內在狀態）。然而（這是論證的關鍵）我們肯定都不會認為，要是行動者不關心別人，那么道德事務基本就無關緊要了。如果羅杰喊道：“獵殺狐貍是道德錯誤！你們的行為是罪惡的！”那么他意圖表達的東西就要求被納入考慮之列，無論獵人們對羅杰的態度如何。當然，獵人們聽到羅杰的話，很有可能還是若無其事地跑了過去，因為他們認為這話是錯的。但即便不同意羅杰，他們也會承認，通過道德判斷，羅杰意圖提出一個重要的實踐考慮。[[36]](#_36_91)這又是反對純粹非認知主義的又一個原因。

那么，道德實踐影響力的源頭究竟是什么？這比明白其源頭不是什么要更容易一些。例如，一個道德規定的適用性通常并不依賴于對不遵循該規定的懲罰。在本章后面的部分，我們將會看到懲罰概念在道德判斷中有非常重要的作用。我們一般認為，犯了道德錯誤的人應得懲罰。但是，認為道德上的不端應得懲罰，和認為道德判斷的適用性依賴于令人不快的懲罰或令人愉悅的獎賞，兩者不能混淆。認為道德判斷的適用性依賴于潛在的懲罰，意思是說，道德命令是一些建議，其有效性在于若不聽從就會引發不想面對的后果。這種實踐建議的重要特征是，如果被建議的一方愿意接受懲罰，或有逃避一般懲罰的特殊辦法，那么“應該”的說法就得要收回。但是，道德判斷看起來并不是這樣。這并不是否認，在試圖向別人解釋為什么不應該（比如）違背承諾時，我們可能首先會提到那樣做會導致的消極后果。我們甚至可能有一種根深蒂固的制度化文化傳統，即通過訴諸某個全能的神的懲罰來支持道德判斷。但是，有些行為可能既在道德上是可鄙的又是不夠明智的，這種可能性不能掩蓋一個真實而重要的差別。當我們說違背承諾在道德上是錯誤的，我們認為這也適用于不顧后果的人、違背承諾但直到去世也沒有受到懲罰的人（暫不考慮死后懲罰的可能性）、很有可能逃避一般懲罰的人，乃至對道德絲毫不以為意的人。這里的“我們”是指泛文化的、人類意義上的“我們”。如果真聽說有某個文化，其成員以實踐表明他們普遍認為犯罪之所以是錯誤的，是因為它們會招致懲罰，而不認為犯罪行為會被懲罰是因為它們是錯誤的，那我會非常驚訝的。

道德判斷的適用性并不來自外部的制裁，這一點可以得到廣泛共識。但有一個歷史悠久的學派認為，道德判斷的適用性來自內在的制裁。按照這種思想，一個犯了道德過錯的人最終在一定程度上傷害了他自己，所以，道德判斷就是類似于教人避免自我傷害的建議。這種觀點有許多不合理之處。首先，雖然經過了幾千年的嘗試，但是這樣的自我傷害到底是怎么回事，目前還沒有人提出一個令人信服的理論。有人可能會簡單地說道，“行惡之人敗壞了他自身的靈魂”，但是在缺乏細節和證據的情況下，這等于什么都沒說。在我看來，這種傳統更多來自于學院派哲學家，而跟日常生活中使用道德判斷的普通人的道德實踐沒有多大關系。考慮以下例子：惡人杰克做了一件明顯的壞事（殘忍傷害了一些他根本不該傷害的人），如果他所屬文化的成員因而都毫不遲疑地斷定他犯下了嚴重的罪行，認為杰克行為的錯誤在于傷害了他自己，這種想法肯定非常古怪。毫無疑問地，折磨的錯誤之處主要是源于強加給他人的傷害！此外，如果杰克的行為足夠惡劣，那么他就應該為自己的行為遭到懲罰。在這個不言自明的真理面前，那種“自我傷害理論”就更說不通了——說一個人自我傷害導致的痛苦不快，使我們有理由（甚至要求）對他造成進一步的傷害，這話是什么意思？跟之前一樣，如果真有一個文化以其道德實踐指出自我傷害是首要的罪行，而對他人的傷害不過是衍生性的，那我會非常震驚。也許大多數道德罪行確實會引起這樣或那樣的自我傷害，但首先，即使排除掉自我傷害（哪怕只是在想象中），道德判斷依然是恰當的，其次，我們對于違背道德規則的自我傷害，會加之以更多的傷害。這兩件事實都表明，道德判斷的適用性并不取決于自我傷害。

既然對行為的道德判斷一般不能等同于建議，這樣看來，似乎道德判斷一般都是定言命令。定言命令是康德的術語，這種實踐命令的有效性不取決于被規定的對象的個人目標［（1783）2002：216］。“關上窗戶”這個命令后面通常還附加著“如果你想要更暖和的話”（但一般都是默認的）。所以要是聽者不想更暖和，那么這個命令式一般就會被撤回。這是假言命令。與此相反，“不要殺害無辜之人”這個道德判斷就不取決于聽者的任何目標。一個人不能以通過訴諸一些稀奇古怪的欲望或目標來逃避一個道德命令。設想一下，面對禁止殺人的要求，某人回應道：“但是我確實享受殺害無辜之人的樂趣啊——其實這正是我人生的主要計劃。”對于這種回應，我們不會撤回道德命令，說“啊……好吧……既然這樣，我想你殺害無辜也沒有什么不應該”。道德命令基本上不依賴于聽者的目標。即使有某種目標是我們的本性（就像康德所認為的）。他聲稱自然母親賦予我們每一個人追求各自幸福的欲望［（1783）2002：217）］，建議我們該如何滿足那個目標的假言命令，還是沒有在我們看來道德命令必需的那種實踐力量，因為道德命令并不僅僅是人們不會通過訴諸特殊目標來逃避的命令，更是人們訴諸特殊目標也不能逃避的命令。

我的意思并不是說所有道德判斷都包含了命令式，或者所有道德命令都是定言命令。我只是認為定言命令在任何道德體系中似乎都至關重要，乃至要是某個陌生的價值體系缺乏這種評價行為的方式，我們就會懷疑這個體系能否算得上是真正的道德體系。康德的定言命令概念有很強的理論背景，也是對猶太—基督教的文化傳統里本來就被闡述很多的一個概念的哲學詮釋。也許有人會抗議說，我賦予定言命令這么高的地位，其實是把某個現代的和西方的道德傳統強加于道德本身。為了打消這種想法，我在此要強調一下，我說定言命令存在于幾乎所有的道德體系中，并不是要把康德的整套體系照搬過來。我對定言命令的理解是極其簡單的：定言命令不是建議該怎樣實現目標的命令。我想大概沒有人會懷疑道德命令通常確實是這樣，而且我還要頗有信心地提出一個人類學的主張，即在所有的文化里都是如此：所有的文化都認同某些傷害他人的行為是錯的，但是沒有任何一個文化會認為，這些行為的錯誤是源于妨礙作惡者實現的目標而造成的傷害。（我在此要鄭重提醒不要認為道德哲學家的觀點代表了他們所屬文化的現實道德信條。）

菲利帕·福特（Philippa Foot）1972年的一篇論文讓人們注意到另一種命令（或者說另一類的“應當”句[[37]](#_37_91)），這種命令似乎恰好介于康德區分的兩種命令之間。禮儀規則（例如，“滿嘴食物時不要與他人說話”）不會被撤回，即使說話人發現聽者毫不在乎禮儀。不管一個人對禮儀的感覺如何，他或她都不應該在滿嘴食物的情況下說話。在這一點上，避免誤讀福特的意思是非常重要的。她不是要支持一種過分保守的觀點，認為我們在任何情況下都必須遵守禮儀。要是你注意到你的晚餐伴侶眼看就要吞下一只黃蜂，而你一時來不及把口里食物咽下去，這時最正確的事情就是馬上提醒她，即便這意味著你要在嘴里塞滿食物時跟她說話。福特的意思是說，禮儀的規則在這種時候并沒有“停止存在”；在某種意義上，“你不應該在滿嘴食物時與他人說話”這個規則仍然是正確的。這一點可以從以下事實看出：事情過后，你可以很自然地說：“為了阻止我朋友被黃蜂叮咬，我必須違反禮儀。”你意識到你破壞了這些規則，這正表明“這些規則”依然存在，因為倘若它們在那種情況下不成立，你根本就不會違反這些規則。

福特小心翼翼地把這些規則稱為“非假言命令”，但是根據我對定言命令的極簡定義，它們就是定言命令。福特不打算把這些規則看作是完整的定言命令，因為它們缺乏定言命令通常充斥著的實踐力量。這里剛好可以借用大衛·伯林克（David Brink 1997）的幾個術語來澄清區別。定言命令（包括制度性規則，如禮儀規則）具有道德的不可逃避性。不管一個人的目標是什么，定言命令都能合法地施加于她身上。但是有許多哲學家（以康德為首）想要道德命令有更多的內涵，即更強的實踐力量，以此在邏輯上區分道德命令和禮儀。他們想要增加的是道德的權威性。我們可以這樣說（此為初步的大體刻畫）：對一個人來說，如果忽視某道德規則是非理性的，或者如果他真有理由去遵守該規則，那么可以說該道德規則具有權威性。我們可以認為道德規范性既包括不可逃避性，也包括權威性。由于缺乏一個合適的詞語描述這兩者的結合，就讓我從此規定，兼具這兩個特征的規范體系擁有實踐影響力。[[38]](#_38_88)（從這里開始，我把這些詞語當成專門術語來用；我大概還應該補充一下，“力量”這個詞在后文依然扮演模糊而通俗的角色。道德具有實踐力量是個簡單的觀察事實；而這種力量是否應該解釋成影響力則是個哲學問題。）

道德權威性在哲學上是個爭議很大的題目（但是反對道德權威性的人常常把他們的反對錯誤地表達成是對道德絕對性的敵意），但我不得不坦承，我很同情康德式的直覺：除了不可逃避性，道德還有某種額外的權威性，而我們的道德言論常常富含這種權威性的意味（而這種權威性能否經得起推敲則完全是另一個問題了）。我們只要想一下定言命令（不可逃避的規范性）可以有多么的軟弱無力，就能引起這種直覺。假設愛達荷州有一些奇怪的異教，相信所有人都應該把自己的頭發染成紫色。我假定這并不只是個建議；它的意思并不是說，我們應該這樣做才能避免惹怒全能的紫蜥蜴神之類的。相反，就像禮儀那樣，它是一套適用于人們的規則，無論他們在乎與否。如果你對這些異教的某個成員說：“我不會把我的頭發染成紫色，因為我根本不在乎你們愚蠢的異教。”他有可能回應道：“這些規則的成立根本不依賴于你在乎與否；你必須要把頭發染成紫色，就這么簡單。”顯然，你還是會（并且也應該）不為所動。你無須否認這個異教確實要求你把頭發染成紫色；只不過這個宗教規定在你實踐慎思中的分量跟你自己對遵守這個規定的興趣成正比：都是零。

在我看來，我們都認為道德要求跟這種規定是不一樣的。沒有任何一個人類文化允許它的成員能這樣輕易地擺脫道德規則的權威性。當然，即便如此，也許道德命令確實只是福特所說的非假言命令（即不可逃避的但沒有權威的命令）的一種，只不過是我們所有人都在“道德異教”里陷得太深，無法意識到這一點（同樣，那個愛達荷州的宗教團體大概也不同意你能夠自主決定擺脫他們的規范體系）。但是要接受這個觀點，代價就是要承認道德的權威性不過是幻覺，承認那些對道德毫不關心的人們其實可以正當地忽略道德，就好像我們可以正當地忽略那些要我們把頭發染紫的異教徒那樣，還要承認如果我們能夠看清事情的本來面目，我們就會意識到人們完全可以合理地嘲弄道，“道德，虛偽的道德！”如果我們確實想要避免這些后果（哪怕這不過是因為我們在一個特定規范框架的浸淫太深），那我們就有理由去試圖解釋道德的影響力是怎么回事。

不論道德規定具有何種實踐力量，這種力量的根源都不在于內在或外在的制裁，不在于某個體制不可違背的規則，也不在于受到道德約定者的欲望或目標。與此相關，在人們日常的觀念里，道德要求跟習俗要求、慎思要求是不同的。（有大量經驗證據表明，即使很小的孩子都會做出這個區分。第四章將討論這些證據。）我倒不是說這個區別簡單明了；我確信這個區分肯定是模糊的，還是多層面的，但它仍然非常重要。讓問題更加復雜的是，很多行為既被道德也被人們的習俗所要求（或禁止）——有的人還可能會認為道德要求自己應該尊重某些習俗框架，例如認為遵守法律是道德上正確的，從而可能進一步認為靠右側開車屬于道德要求。因此，在很多情況下，這兩者很難被分開。

上文承認，道德其實可能只是人類的習俗；問題是（當我們“置身”道德之中時）我們不是這樣看待道德的；我們覺得道德具有某種超越習俗的實踐力量。這種力量該如何解釋，這是一個可以追溯至柏拉圖的哲學問題，當然還可以更早。目前我們并不是試圖為這種實踐力量辯護，或為它提供理論說明。我們只是簡單承認道德判斷體現了這種力量，以期能夠理解這些判斷的基本性質。在本書較后的部分，我們會再回到道德權威性的問題，探討它能否被納入自然主義的框架中。

## 2.5　道德判斷的主題

道德判斷的“主題”，我指的是道德謂詞能夠被合理地加于其上的那些事物。這有什么限制嗎？一個可能的限制是：一個價值體系要稱得上是道德體系，某些特定的謂詞必須要被應用到某些特定的主體上。比如也許會有人認為，任何不認同“殺嬰是錯誤的”的價值體系不能被稱為一個道德體系。一種可能更弱的約束是：一個價值體系要稱得上是道德體系，其中的謂詞必須用于某些特定類型的對象。例如，一個價值系統如果只認可有關如何穿鞋的要求，就不能算作道德框架。菲利帕·福特曾經主張，只是因為某人以一定的方向繞著樹跑步，或者在月光下觀看刺猬，就把他視作壞人，這樣的評價就不是出自道德的角度；這種事情就不屬于道德敗壞的事情的類型（Foot 1958：512）。反對這一觀點的是，人類學家發現有些很奇怪的事情似乎在道德上遭到譴責：馬來西亞的舍蔓人（Semang）認為，在暴風雨的時候梳頭或是在早晨扔長矛，都是罪惡的行為（Murdock 1980：89）。

我對第一種限制的存在沒有什么信心，因為涉及道德的事情大都乃至全部存有爭議。如果說禁止殺嬰是所有道德體系的必要特征，這就排除了對是否允許殺嬰進行道德爭論的可能性。我們的論證不能太狹隘；在整個歷史進程中，許多群體都認為，至少在特定環境下，殺嬰行為是可以接受的（有時甚至是必須的）。我不是說這樣的文化觀念跟我們的文化一樣是“正當的”；也許任何允許殺嬰的觀念都是錯誤且有害的，這跟我說的也不矛盾。我的意思只是說，“允許殺嬰”至少可以算作錯誤和有害的道德觀念。如果只是為了強調自身的道德觀點（即殺嬰行為是不可容忍的）而堅持認為殺嬰行為的價值體系甚至都算不上是“道德體系”，這樣做只不過是妨礙我們對這個問題的理解（盡管如此，我并不完全排除存在這類具體限制的可能，但是我可不知道它們到底是什么）。

對于第二種限制，我想更大膽一些。我們注意到，道德謂詞通常應用于某一類事物上。首先，道德謂詞最自然地被用于行為和人。我們既在道德上評價特定的行為（“你上周二對你母親的撒謊行為是卑劣的”），也評價一般的行為類型（“為了自己的利益而撒謊，這種行為是錯誤的”）。此外，我們似乎也評價一個人和一類人，盡管當我們評價一類人的時候（“所有說謊者都是邪惡的”），這個評價針對的大概還是性格特征或行為類型。相對而言可能更有爭議的是我們有時似乎也對事態進行評價：“沒有奴隸制的世界比有奴隸制的世界更有道德。”我并不認為這三種范疇窮盡了所有道德評價的對象。我們有時會把一個信念描述為有害的，或把一件藝術作品看作是道德墮落的。我們也會把制度、法令或整個社會看作是道德敗壞的。另外，在某些文化里有類似“道德污染”的概念：這指的是普通物體或地方所具有的一種性質，任何接觸過這些的人都會感染上這種性質。

盡管道德謂詞能形容的東西有很多，但我們還是做出一定的概括。一系列詳盡的跨文化研究在許多道德體系中都發現了大致的共同點：（1）負面評價某些傷害他人的行為；（2）看重互惠和公正；（3）要求人們的行為與其在社會等級中的地位相符；以及（4）在事關身體的規定中（例如女性的月事、飲食、沐浴、性事以及對尸體的處理等），純潔或污穢的概念通常處于主導地位（相關的討論和參考文獻，參見Haidt and Joseph 2004）。前三類范疇都涉及人際關系，我們從中可以得出以下結論：所有道德體系的大部分內容都是用來保護和維持社會秩序、解決人際關系沖突和遏制過度追求個人利益的規范和價值。道德領域尤其有很多東西跟可能對別人造成彼此傷害有關。（這個觀點的經驗證據，可參見Nichols 2004第七章。）看起來道德的目標是服務社會。第四個共同特征似乎是關于自我導向的行為，但這可能只是表面現象而已，因為我們應該記住，西方文化里屬于自我導向的行為，可能從其他文化的觀點來看則不然。（換句話說，就像把狹隘的道德領域觀念投射到其他文化是錯誤的一樣，把狹隘的傷害或自我導向觀念投射到其他文化也是錯誤的[[39]](#_39_84)。）這并不是否認人們確實會做出純粹的自我導向的道德判斷（比如，用國旗清洗衛生間是錯誤的[[40]](#_40_82)），但是說任何道德體系的大部分內容是關于人與人的關系，應該也沒什么問題。

我這樣說并不是對任何道德體系的必要特征進行概念分析，只是做出一項觀察而已。也許魯濱孫可以創造一整套命令體系，全都是自我導向的（例如，“去島的南邊是被禁止的”等），并且憑借這些命令特有的權威性，它們仍然是道德命令。但或許有人會認為，魯濱孫創造的道德體系都不能算是“道德”，正是因為它們缺乏他人導向的內容。坦白講，在這個問題上，兩者我都可以接受；我很懷疑英語的“道德”一詞的意義是否足夠明確從而能給出唯一的答案。但是，只需要注意到魯濱孫創造的這個自我導向的“道德”體系是多么怪異，我們還是可以得出一個有用的結論。支配人際關系的道德似乎是自然且必需的，而與人際關系無關的“道德”（如果這種情況是可能的）則是古怪和亟須進一步解釋的。我們感興趣的是“人類道德是自然選擇的產物”這個假設，而不是“任何道德都必定是自然選擇的產物”這個極度不合理的假設。因此，人際關系是人類道德的核心對象這一觀察依然是有價值的證據，并沒有受到如下事實的影響（如果這是一個事實的話）：道德的概念邊界過于模糊，乃至某些假想的“道德”可以有完全不同的主題。

## 2.6　應得

如果維護社會秩序是道德的重要目標之一，那么我們可以預料到，一個人倘若違背了道德要求，就會招致其他社會成員對他的負面回應，他們以此鞏固道德規范。我們還可以估計，道德判斷這個概念本身，同違背道德的行為和該行為導致的負面回應之間的特定關系（以及遵循道德的行為與該行為導致的贊揚之間的特定關系）密不可分。我認為，在我們眼中這種關系確實是道德的關鍵：我們認為，某些行為應得某些回應。我們對應得這一概念太過熟悉，以至很難拉開距離，認識到這個概念有多么奇特；它看上去是如此簡單明了，以至于我們被蒙蔽了也不自知。

假設有一個由社會生物組成的群落，他們采用的命令大部分是用來支配人際關系的（“不能為了個人利益而盜竊”，“遵守你的承諾”，等等），此外，他們還似乎把這些命令看作是定言命令：不管道德規定所施加的人是否關心，也不管對方的愿望是什么，他們都堅持這些命令，還要求成員們必須要將其納入實踐考慮。到這一步為止，這些生物都跟我們一樣。但這個社會跟我們有一點不同。在這個社會里，即便有人未能遵循規定（比如有人為了出于自私而違背承諾），也不會遭到同胞的批評。這不是因為他們不理解批評的含義：他們會批評彼此實踐生活中的愚蠢行為，例如把鑰匙鎖在車里或睡前喝太多咖啡等。但是，他們不認為批評是那些情況所要求的。我們可以說這個想象中的社群根本不具有應得的概念。當然，要是有人以極端的方式打破了某條社會規則（比如說殺害許多同胞），那么，為了社會秩序，大家也許會決定把他關起來，或者也許會在殺人與被囚禁的痛苦之間建立一定的比例關系，以此來阻止潛在的違法者。出于這些理由而執行的處罰，可能是個“好辦法”，可能很有效果，但并不是錯誤行為所要求的。被懲罰的人做了令其他成員討厭的事情，他們也許還因此討厭這個人，所以把他監禁起來在一定程度上令他們感到滿足——但是他們心里想的一直都不是：這正是他應得的。

顯而易見，我們和他們的差別可不止一點半點。我們假定他們沒有應得的概念，但這似乎意味著他們缺乏的其實遠不止此，雖然究竟還缺乏什么不是很容易準確估計。其中一個主要的方面是，這個社會中的成員肯定缺乏正義觀念的核心要素：正義體現為人們得到應得的賞罰。如果告訴這些生物，那個被普遍認為有罪而當眾受懲罰的人其實是無辜的，他們也不會感到有什么困擾。倘若我們抗議道，“但是這樣做是不正義的；他不應得到這樣的對待！”他們大概壓根不知道我們在說什么。此外，由于缺乏對應得的認識，他們也就沒有對負罪感的認識——負罪感不就是一種包含“自己的行為應得懲罰”的判斷的情感嗎？（下一章會對此展開論述。）沒有負罪感，這些生物也就不可能有我們說的“道德良心”。我們甚至還可以猜想，他們也不會真正欣賞典型的好萊塢電影和簡·奧斯汀的小說，因為這些電影和小說的結局都是正義的賞罰得以實現。他們是無法從中得到任何滿足的。

既然這個想象社群的成員不能理解這些概念，那么我相信這會給之前的提議，即他們做出的是道德判斷，打上一個大問號。沒有了應得、正義、負罪感和良心等概念，他們的定言命令對我們而言一下子就顯得非常陌生（就算這些道德命令表面上具有權威性，而且對成員間的關系很重要）。在一個缺乏這些概念的思維框架里，人們發出的“不許殺生”的命令，到底有多少力量呢？或許這有點像我們命令動物“不許殺人”。假設你聽說有鱷魚殺了人，逃走后再沒有出現過。也許有證據證實它此后再沒有傷害過人類。再想象一下，如果故事中殺人的不是鱷魚，而是一個人，他隨后也逃離現場，再沒有出現過。比較一下你聽完這兩個故事后的感覺。我們設想的那些生物缺乏應得概念，因而不會感到這兩者有什么區別；他們也不能理解到逃脫懲罰的人和僅僅是逃走的鱷魚之間的區別。

仔細審視我們日常的應得和正義的概念就可以發現，我們眼里的世界似乎具有某種“道德平衡”。惡行打破了這種平衡，而適當的懲罰則是恢復平衡的手段。修復平衡令我們滿意，但要是惡人逍遙法外，我們也相應地心存不安（參見Lerner 1980；Vidmar and Miller 1980）。小說里的惡人最終受到應得的痛苦懲罰，這帶來的滿足很少有其他的小說主題可以相比。我們的滿足感并不是因為“除掉禍害”，也不是因為惡棍的悲慘結局給其他潛在的惡棍留下了教訓。挑明了說：我們高興就是因為這混蛋活該落得如此下場！我們現在所關心的并不是“平衡”這個概念能否說得通。我承認，要是仔細推敲這個觀念（尤其還不想引入超自然的力量），它聽起來確實非常神秘。但是，我們現在并不是要攻擊這個觀念，或為之辯護；我們現在只不過是試圖理解乃至“感受”人類道德的特殊之處。

這里的觀點可以用一種更強的方式表達：對某人應該做某事所下的道德判斷，包含了她不做的話就應得某種形式的懲罰的意思。應得的懲罰并不一定是正式的刑罰，像坐牢或罰款之類的，因為道德錯誤可以源自比較細微的行為（例如踩到別人的腳而不道歉），但是也許我們覺得，輕微的道德冒犯也應得到他人的批評，或者哪怕只是冷眼相待。對于需要什么樣的負面回應，我們可能沒有明晰的想法，但我們還是隱約覺得冒犯者“應該遭報應”。此外，我們可能認定總的來說不懲罰更好，但同時仍然判斷某些負面后果是應得的。甚至有時候，我們覺得應該“把另一邊臉伸過去，隱忍不發”，但與此同時我們又相信超自然的神明會在審判日給對方相應的懲罰。

但是這種觀點也許太強了。或許我們能找到一些道德判斷的例子，其中某件事被判斷是錯誤的，但沒有人認為犯錯者應受到任何負面的影響。也許足夠輕微的小錯就是如此。可是，我們仍然可以說，這些道德判斷只有在某種特定的概念框架里才有意義，而這種框架需要能夠辨別相對嚴重的錯誤行為，即需要懲罰的錯誤。有沒有這樣的“道德”體系，其中所有的錯誤行為都只不過是小錯，沒有什么應得明確回應，而且根本沒有應得的概念？對此我深表懷疑。

在討論應得概念的過程中，負罪感的話題出現了幾次。第三章對負罪感有更詳細的討論，但由于負罪感與應得概念之間的重要聯系，我想在這里作一點簡述。這種聯系在于，應得和負罪感都是補償或修復的概念核心。一個人做了錯事，如果接受了應得的懲罰，那么平衡就能夠得以恢復。如果她經歷了適度的悔恨，可以想見負罪感在其中起到了主要作用，這樣一來平衡也能得以恢復。事實上，犯錯的人被懲罰卻沒有悔恨，這會令人感到不安。要是罪犯服完刑回到社會，卻完全對先前的惡行毫無懺悔之意，哪怕他沒有任何再犯罪的意圖，我們對此也會有憂慮。我們接受他已經為自己對社會的損害付出了代價，但盡管如此，我們還是會擔心懲罰沒有真正“起到作用”。這種傾向的發展可以追溯到童年時代。讓小孩子評價違背道德者的臉有多壞，大多數6歲和所有8歲的孩子都認為笑臉比悲傷的臉更壞，而4歲的孩子則不會做出這個區別（Nunner-Winkler and Sodian 1988）。適當的負罪感（悲傷的臉）是我們道德發展的重要一步。對于犯下了普通道德錯誤的人，我們認為只要他們有負罪感，這就足以補償他們的過錯。在我們看來，一個不會產生負罪感的人可能是危險的，并肯定是有缺陷的，就算他的實際行為似乎并沒有什么道德污點。一個人要是不能產生負罪感，那么也就缺乏一個“內在的自我懲罰系統”——我們認為這種系統是可靠地引導道德行為的重要機制。

## 2.7　總結與預告

本章開篇就揭示了，要在生物意義上理解人類如何被設計成具有諸如愛、利他主義、同情等親社會性情感是比較簡單的：用親緣選擇就足夠解釋了（這并不是否認這些情感也可以通過其他方式演化而來）。那些論述使我們注意到，這樣的解釋框架其實遠遠不足以解釋做出道德判斷的能力。這使得我們直接面對這個問題：“到底什么是道德判斷？”這一章接下來的部分則強調人類道德判斷的幾個重要特征。下面是對這些特征的總結：

——道德判斷（作為公共言論）通常表達了意動態度，例如贊同、意圖等，或更一般地說，是表達了對標準的認同；但道德判斷同時也是信念的表達，即它們是斷言（assertion）。

——無論行為者的利益/目標是什么，有關行為的道德判斷據說都要構成實踐慎思的考慮因素；因此，它們不屬于建議。

——道德判斷據稱是不可逃避的；沒有“退出”道德約束的途徑。

——道德判斷據稱是超越于人類習俗之上的。

——道德判斷的核心是調控人與人之間的關系；道德判斷的目標似乎尤其是抵抗蔓延的個人主義。

——道德判斷蘊含著應得和正義（一個“賞罰”系統）的概念。

——對于我們這樣的生物來說，負罪感（或“道德良心”）是規范道德行為的一個重要機制。

需要注意，這份清單包含了看待道德的兩種方式：一種是從道德判斷的獨特對象（人際關系）來看待道德；另一種則是從道德的“規范性形式”（一種極有權威性的評價方式）來看待。這兩種特征各有應得的地位。任何一種有關人類道德能力演化的假設，要是不能解釋自然選擇為何青睞兼具這兩種特征的判斷，就肯定是不完整的。

也許有人會覺得，清單里面的有些特征只是從對人類道德的特征觀察得來的，但另一些則很可能是有關道德判斷本質的概念真理。但大體上，本書不需要處理這個微妙的哲學區分，我也打算放下這個問題。因此，我并不是宣稱這個清單成功囊括了道德判斷必要而充分的條件。我們對道德判斷的概念還不大可能清晰明確到能讓我們列出這種條件的程度。更明智而謹慎的說法是，只要一個價值體系滿足足夠的上述條件，就能夠稱得上是個道德體系。一個更大膽的說法則是，清單中有些條件（至少有一個，但并非全部）是道德判斷的必要特征，要成為道德判斷，至少還要符合足夠的其他條件。但無論哪種說法正確，多少才算是“足夠的”？強制規定是毫無意義的。事情取決于我們這個擁有語言能力的群體，在遇到另一個不熟悉的社群時，對于他們是否有道德體系這一問題會做出怎樣的實際決定：如果他們具有跟我們不同的價值體系，該體系滿足以上清單的四個條件，在他們的語言中，有一個詞語（比如“woogle價值”）專指這個體系，那么，我們可以把“woogle”翻譯成“道德”嗎？我完全不知道這種反事實的決定會是怎樣。我的意思只是說，在這個決定中，上述各條件都是重要的考慮因素（在6.4節我們將會再次遇到這個議題）。

這個清單有可能缺少了一些重要特征。有的可能是我疏忽了；其他的則是我故意省去了。例如，有人論證說道德判斷獨有的特點在于它是普適的規定（參見Hare 1952，1963）。我一向認為這個斷言是有問題的，無論它是概念分析還是經驗觀察。讓我們的道德規定能夠普適化，這或許是個值得追求的目標，但是這和說道德規定必須要能普適化，否則就不能算作“道德”規定，完全是兩碼事。很多社群的價值體系，我們會毫不遲疑地稱其為“道德的”體系，但是這些體系容許個別的判斷。例如，很多（如果不是大多數的話）道德體系都允許截然區分本社群成員和外人。這個區分可以有普適判斷的支持（“任何人對他或她的社群具有特殊的責任”），但是在現實世界里，個別判斷似乎不太可能是從這種普適的基礎推論出來的。雅諾瑪人（Yanomamö）[[41]](#_41_82)曾經認為殺戮遇到的外國人沒什么關系，他們因此而惡名昭著。誰能說他們就是把“外國人”這個詞看作是普適的范疇？或許對他們而言，外國人就是雅諾瑪族以外的所有人，或是任何沒有皮日波利瓦[[42]](#_42_76)血統的人。我不知道他們把“外國人”理解為共相（一般概念），還是無法消除的殊相，但這不意味著在他們的規則“殺害外國人是可以的”是否算是道德自由的問題上，我就相應地也要持中立態度。（對于所謂的愛國美德和普適的道德準則是否有矛盾，我也有些顧慮。）

現在我們對什么是道德判斷有了些許認識——或者至少對本書如何理解這個術語有了一定的了解（我希望兩者是同一回事），于是就可以轉向我們真正的任務：人類做出這種判斷的傾向是生物自然選擇的產物嗎？有人可能想回答“不是”，因為我們已經認識到道德判斷是深有權威的親社會規定，而他可能覺得生物自然選擇青睞只是自私自利的行為。但是第一章已經充分證明了這種觀點是錯誤的；自然選擇在一些環境下完全可以偏向親社會性狀。我們現在感興趣的假說是：進行道德判斷的傾向就是人類的親社會機制之一。

還有一個問題沒有解決：道德感是何時演化出來的？在第三章中，我將論證語言的出現是道德概念與某些道德情感出現的先驗條件。道德感為什么得以演化？第四章將論證，在個體與人際這兩個層面上，道德判斷作為承諾，能夠促進做出道德判斷的人的繁殖利益。道德感是如何演化的？第四章將探討一些經驗證據，這些證據支持以下假說：自然選擇通過修正情感來形成人類的道德感。此外，第四章還將進一步推測，人類進行道德判斷的能力的核心在于人類能夠“投射情感”到外部世界。該章的結尾將回顧一些經驗證據（多數來自發展心理學），表明本書到目前為止的主要假說，即人類的道德感確實是生物自然選擇的產物，是值得認真對待的，而不是一個“不過如此”的故事。第五章和第六章可以作為本書的第二部分。它們處理的問題是：我們從本書第一部分提出的假說中可以推導出什么元倫理學的結論？

# 第三章　道德語言與道德情感

## 3.1　動物的規范性（但非道德性）心智生活

我小的時候，家里的狗大部分時間都關在后院里。那只狗總想著要掙脫束縛，跑進附近社區的熱鬧世界，雖然它知道這是不被允許的。它會慢慢地踱向柵欄，像是在研究一種有趣的氣味，同時密切留意著窗戶那邊，以確保沒有人留意它。一旦覺得沒人監視，它就會跑到柵欄跟前，用盡全力地挖土和推柵欄。如果逃跑還沒成功就被發現了，它會繼續先前的偽裝，仿佛在研究什么有趣氣味，雖然表面若無其事的樣子，但還是忐忑不安。但要是在離自由還差幾英寸的時候被發現，它會先焦慮地往后瞟一眼，迅速計算來人還要多久才能穿過后院，然后加快推動柵欄（最后經常成功脫逃）。

在某種意義上，這完全是我童年的真實故事；我的狗過去就是這么干的。但在另一種意義上（在任何非哲學的語境下堅持這種意義都是乏味的），這并不是真的。嚴格來說，我并不相信我的狗“知道它不準出去”或者真的感到“負罪感”。它確實知道它會受到懲罰，也不想被懲罰。它把家里的規定內化了，但它沒有把規定作為規定來理解。雖然它的心智在許多方面表現出令人驚訝的狡黠，并且聰明到可以對“走”和“坐”這樣的命令有所回應，但它還沒有能力去形成“離開后院，或者別的事情，是被禁止的”想法。

本章的目的就是支持這個論斷。我將首先探討動物可能具有什么類型的規范性思想，然后再論證它們沒有道德思想。該論證的基礎是如下觀點：語言是擁有道德概念的先決條件。隨后我會對道德情感展開論述，主要的意圖是表明某些情感具有豐富的概念意涵。最后的結論是：語言是擁有某些道德情感（其中最顯要的就是負罪感）的先決條件。這可以告訴我們道德意義是何時演化出來的——這不是指年代意義上的時間，而是指道德感與其他性狀出現的先后關系。

我開頭舉的是狗的例子，但集中討論猩猩這個更復雜的例子可能更富有成效，因為我們可以合理地認為，猩猩在很多方面類似于人類發展的某個階段。因此，討論猩猩在道德領域具有或缺乏什么，就有希望幫助闡明在人科動物演化的道路上，道德判斷能力是什么時候形成的。[[43]](#_43_76)有些學者宣稱，猩猩無可置疑的社會復雜程度，以及貌似遵守規則的行為，就足以證明猩猩具有道德感（Sapontzis 1987；Harnden-Warwick 1997；Waller 1997；同時參見Gruen 2002）。我將論證這種觀點是錯誤的。我不認為我捍衛的只是少數人的看法。我估計大多數人都愿意同意猩猩沒有進行道德判斷的能力，但是至于它們為什么沒有這種能力，卻很少有人能給出一個不那么模糊的答案。

為了行文清晰，我們首先來區分關于動物與道德之間關系的三個可能的問題：第一個問題是，動物是不是道德主體——它們是否具有一定的本性，從而能夠對我們提出道德要求。雖然這是個非常重要的實踐問題，但并不是我們這里所感興趣的；我提出這個問題僅僅是為了把他同別的問題區別開來。第二個問題是，動物是不是道德行動者——即它們的本質是否使得其行為可以在道德上受到責備或稱贊。雖然在日常用語中，我們經常使用道德評價來描述動物（“惡狗！”），但極少有人是在字面上理解這種評價。需要注意，我這里說的是道德評價。根據某些能具體表述的標準來評價動物（比如，一匹“好賽馬”或一只在行為意義上守規矩的“好狗”）并不成問題。第三個問題（這才是我們感興趣的問題）是，動物本身是否能做出道德判斷。這同第二個問題并非不相關（我在“導論”中也提到過），因為很可能有人會論證說，只有當一種生物能夠區別是與非時，他的行為才能被評價為在道德上值得贊許或責備，而區別是非的前提就是道德判斷能力。我懷疑兩者之間的聯系雖然在直觀上像是那么回事，但實際上并不是成立的（參見2.2節），但我隨后構建的論證跟這種聯系無關。我更傾向于直接面向第三個問題。

靈長類動物學家弗蘭斯·德瓦爾（Frans de Waal）對描述性社會規則與規定性社會規則所做出的區分（1992；以下的頁碼均出自這篇文章），對于我們的討論是個有用的起點。描述性的社會規則只是生物體對同類的反應規律，這種意義上的規則甚至也可以適用于魚類和昆蟲。相對而言，規定性的社會規則不僅僅是被單純遵循的規律，而是個體“因為他人積極的強化作用而遵守”的規則（244）。談到“避開母親因為保護孩子而產生的敵意”這個規則，德瓦爾說，“當群體的成員開始意識到它們自己的行為和這個母親的行為之間的聯系時，于是以盡可能降低相應的負面后果為目的時”，這個規則就變成規定性的（同上）。德瓦爾認為，這些規定性的規則是道德系統存在的必要但非充分條件。它們（看起來）確實足以產生德瓦爾所說的“社會規律意識”。他對這種意識的定義是“有關自己（或其他成員）應該受到怎樣的對待，和有關資源應該如何分配的一組預期。如果有個體違背這些預期，進而造成了對自己（或其他成員）的不利影響，就會招來負面的回應：通常而言，處于從屬地位的個體會反抗，處于統治地位的個體則會懲罰違反者”。（德瓦爾認為，在這個定義中附加的“其他成員”大多數指親緣成員。）德瓦爾確信猩猩具有這種“意識”，但是他隨即補充道，猩猩區分“被接受與不被接受的行為”的能力，在什么程度上能夠被視為它們“意識到‘善’與‘惡’，‘公正’與‘不公正’”的證據，“還是不清楚的”（254）。他寫道，類人猿沒有口頭語言，所以缺乏這種“形成和交流抽象規則的認知”能力（241）。因此德瓦爾承認類人猿具有道德性的“前身”或“基本組成單元”，但不承認它們有真正的道德（同時參見Waal 1996；Flack and de Waal 2001）。

在我看來，上述“社會規律意識”的定義有些含糊其辭。它的模糊之處在于“期待這樣或那樣的事情應該發生”的可能含義。我們都知道，“應該”在不同的語境下有非常不同的用法。這里有三個截然不同的用法：

（i）明天應該要下雨。[[44]](#_44_74)

（ii）如果你想明天早點起床，你就應該早點睡覺。

（iii）一個人應該避免殺害無辜的人。

這三個例子都可以很容易地用“應當”取代“應該”。第一句話的用法屬于“預言性的‘應該’（應當）”，是指證據顯示某件事情發生的可能性。電視天氣預報員每晚都使用這種意義的“應當”。因為任何具有最基本的關聯學習能力的動物都能預期將要發生的事情，所以任何這樣的動物都可以對應該發生什么持有某種期待。這里的“應該”沒有添加什么實質內容，因為說“X期待這樣或那樣的事件應該要發生”（預言性的“應該”）和說“X期待這樣或那樣的事件將要發生”并沒有什么不同。另一方面，如果我們把“應該”一詞解讀為更明顯地具有評價意義的詞語（比如道德意義上的“應該”），那么自然的說法則是“X判定這樣那樣的事件應該發生”。換句話說，一個人要么預期某事將要發生，要么判定某事應該發生；但是仔細推敲一番，就會發現所謂一個人預期某事應該發生，確實是非常古怪的觀念。

這些學究式的考察不是要批評德瓦爾的觀點，而是為了預先提醒讀者，動物具有某些包含“應當”或“應該”等內容的心靈狀態也許是可能的，但這種想法完全無助于證明我們可以合理地認為動物具有道德判斷。有的人可能會論證說，動物不可能形成包含道德“應該”的信念，因為它們根本毫無信念可言。但這不是我的策略。在下文中，我將會十分謹慎地避免提出關于動物認知的任何特定理論，因為動物認知是個很有爭議的題目，本書在這個題目上采取特定立場只會損害我的論據。面對潛在的反對者，我愿意接受一種開放的觀點，承認猩猩和其他動物都可以算作具有各種信念，不過我將會論證，它們仍然不具有與道德問題相關的信念。

假設預期是某種信念狀態，我們之前已經看到，動物可以有某些包含預言性“應當”（“ought”of prediction）的信念。那它們也有包含假言性“應當”（hypothetical“ought”）的信念嗎？答案要是肯定的，那就很有意思了，因為這樣的信念更能配得上“規范的”這個標簽，而那些包含預言性“應當”的信念則肯定不能算是規范性的信念。正如我們在第二章提到的，假言命令——例如前文的例句（ii）——這種規定的正當性依賴于主體的某個目的［Kant（1783）2002：216］。如果該主體其實并沒有那個目的，那么“應當”的主張就得撤銷。因此對于（ii）而言，如果對方打算一直睡到明天中午，那么“你應該早點睡覺”這個命令就必須被撤銷。假設吉姆相信（1）莫莉想要早點起床，還相信（2）早點睡覺對莫莉來說是實現其愿望的最好方法。那么，現在的問題就是：吉姆是否相信（3）莫莉應該早點睡覺。答案取決于我們如何在一個重要的區分中抉擇。一方面，可能（3）無非就是（1）和（2）的邏輯結果，這樣一來，雖然吉姆應該相信（3），但他可能實際上不相信（因為人們經常不相信他們信念的邏輯后果[[45]](#_45_74)）。另一方面，或許相信（1）和（2）構成相信（3），所以既然吉姆相信（1）和（2），他就已經相信（3）了。我們之前見過類似的例子：如果有人相信明天可能會下雨，我們可以合理地說他相信明天應該下雨（這里是指預言性的“應該”）。后者不是隨附前者的信念，也不是由前者推論出的結果；相反，它們只不過是描述同一個信念的兩種方式。我不得不承認，對于（1）+（2）與（3）的關系，我并不清楚哪種正確。

如果猩猩不能對別的個體有什么欲望形成信念，那么上面這些討論都沒有什么意義。然而，有一些證據（雖然確實有爭議）顯示它們可以對其他猩猩的心靈狀態形成高階信念（Premack and Woodruff 1978；Premack 1984；O’Connell 1995）。此外，它們看起來有也能力對如何最好地實現欲望形成信念。因此，猩猩可能具有上文（1）和（2）這種類型的信念。例如，一只名為佛洛德的猩猩看著一只年輕的母猩猩正吃力地去夠枝上的食物，也許他知道這只母猩猩想得到食物，還知道它得到食物的唯一方式是用長棍子把食物捅下來。那么，如果（1）+（2）與（3）之間的關系是構成性的，我們可以把“那只年輕的猩猩應該使用棍子”這個信念灌輸給佛洛德。但如果這種關系不是構成性的，那么我們就得不出這個結論。我們并不真的需要通過長篇大論來試圖證明其中任何一個結果；為了論證的需要（哪怕這只是對那些想自由地把信念賦予動物的人們做出的讓步），我們允許猩猩可以被賦予包含假言“應當”的信念，而這種信念，我們也可以承認，的確可以被當成某種規范性判斷。我真正想要證明的觀點是，即使我們做出這么大的讓步，把具有規范性的心理內容灌輸給猩猩（而我并不是在論證我們應該這樣做），但是這仍然離賦予猩猩任何類似道德判斷的信念相差甚遠。

## 3.2　猩猩沒有道德判斷

在描述猩猩對規定性社會規則的運用時，德瓦爾的措辭非常謹慎。他寫道，猩猩能夠意識到“被接受與不被接受的”行為。這與意識到可接受的與不可被接受的行為是非常不同的：前者僅僅意味著知道這些行為將要遭到敵意，而后者則意味著判斷這些行為應得敵意。（后綴“able”常常不能從字面上理解。“可欲求的”并不意味著“有能力去欲求”，而是“值得去欲求”。有的人渴望發生種族屠殺這個事實證明了種族屠殺可以被欲求；但是承認這一點很難令我們同意屠殺是值得欲求的。）上文論證了，值得和應得是道德判斷的核心。沒有這些東西，也就沒有什么正義感、負罪感和道德良心；沒有它們，“道德主義攻勢”只能是一個比喻。事實上，一個嚴謹固執的人可以合理地堅持，要是沒有應得的概念，我們就應該避免使用“懲罰”，而改用更為中性的語匯，例如“負面回應”。

我們能合理地認為猩猩有“某些行為值得特定的回應”這種信念嗎？讓我們來看看猩猩是如何用“道德主義攻勢”（我們要知道在這個出自Trivers 1971的短語里，“道德主義”純粹是用作比喻）來抵觸錯誤的行為。公猩猩和母猩猩打斗的時候，前者一般會避免用到能造成相當傷害的巨大犬齒。人們觀察到，在極少數的情況下，公猩猩用它危險的犬齒來對付母猩猩，受害者抗議的聲音就會發生變化——這種變化使得整個群落都對這只雄猩猩發出譴責的咆哮聲，有時整群母猩猩還會聯合起來趕跑攻擊者（de Waal 1992：247）。這不正是表明群落成員判斷雄猩猩這樣使用犬齒應得的特定懲罰嗎？因此這不就包含有道德判斷嗎？我認為不然，因為我相信只要把厭惡、抑制和欲望賦予猩猩，我們就完全能夠以此解釋它們的所有行為——沒有必要把違規、禁律或應得的懲罰或別的什么道德概念賦予猩猩。此外，我認為，我們還有理由不該把這些道德概念賦予猩猩。我將先提出兩個不準確的解答，再簡單地提出更好的解答；隨后我將給出自己在這個問題上多少有些異類的看法。

首先，人們可能覺得道德概念過于抽象，因此認為動物不能有這些概念；有人可能認為語言是抽象概念的必要條件。但是，一旦我們同意動物有信念，看來就無法避免承認動物的信念具有抽象的內容。即使某些刺激是全新的（Premack 1983），猩猩也能夠學會區分相同與不同的刺激，而且豬也有這種能力（Keddy-Hector et al.，forthcoming）。因此，有人論證道，這就使我們能夠把包含了非感覺性的關系，即同一與差別這對抽象概念的信念灌輸給動物。猴子似乎可以認識到一般的母子關系，而這很顯然地允許我們把包含了某某的母親這種關系性質的信念賦予它們（Dasser 1988；Cheney and Seyfarth 1990）。達爾文注意到，狗能夠運用狗的大體概念，因為“當一只狗見到另一只狗在遠處，顯然它看到的是抽象意義上的狗；因為當它走近對方時，要是發現原來是它的朋友，它整個身體姿態就會突然發生變化”［（1879）2004：105］。[[46]](#_46_72)還有很多其他的例子表明語言并不是擁有抽象信念的必要條件。

人們不愿意把道德判斷賦予動物（或者一般而言的非語言使用者），另一個可能的原因是人們隱隱約約地感到道德概念過于復雜。我覺得我對于“復雜性”的理解差不多也就只是直觀感受而已。禁止的概念真的比香蕉的概念更復雜嗎？在什么方面是復雜的？無論如何，到底語言的什么性質引發了概念的復雜性，這一點還不清楚。我（直觀上）感到與……一樣這個關系概念非常復雜，然而一些非語言使用者似乎具有這個概念。

如果既不是因為復雜性，也不是因為道德的抽象本質，我們才不該把道德賦予猩猩等動物，那么到底是因為什么？這常被忽視的問題有幾個可能（它們之間可能是互補的）的答案。馬克·豪瑟（Marc Hauser 2000，第9章）論證道，動物缺乏道德概念與道德情感，原因在于它們缺乏自我意識（這不同于自我識別。用鏡子做的實驗證明動物有一定的自我識別能力）。杰弗里·塞爾—麥柯德（Geoffrey Sayre-McCord，即將展開的私下討論）論證道，要把事物看作是更好的或更壞的，動物必須使用一些概念，這些概念“必須不僅僅能反映事物是否滿足一定的標準，而且也能反映這個標準本身是否得到了證成（大意是說這個標準本身滿足相關的價值標準，并且把它作為標準的來使用也是因為它通過了這項考查）”。要運用一個規范性概念，僅僅是根據一定的標準來證成個人行為是不夠的；使用者還必須能對該標準本身是否得到證成足夠敏感。一旦發現了符合這個概念的事情其實并沒有為進行或避免行為提供理由，那么使用者就要愿意修正這個概念的使用標準。既然動物缺乏這樣靈活的復雜認知（雖然塞爾—麥柯德并沒有明確斷言，但這樣假設似乎比較妥當），因此我們必須得出結論，它們并沒有對它們的世界或對彼此做出真正的評價。兩個論證都很有道理。但是我自己的簡單理論非常不同，而且過去的經驗告訴我，它在哲學家中并不受歡迎。然而，寫本通篇都明顯正確的書是非常枯燥的，因此，我想請讀者容忍我用一些篇幅來簡述對于為什么沒有語言的動物不做道德判斷的一種另類分析。（這些內容可以同本書的主要論證思路分割開來，所以如果讀者愿意承認“語言是進行道德判斷的前提”這種廣為接受的觀點，那么省略這部分內容，直接跳到3.4節也沒有關系。）

首先可以注意到，有許多詞匯我們都不會用來描述動物的信念，哪怕我們傾向于慷慨地給別的生物賦予信念。無論羚羊多么不喜歡獅子，我們都不會真的認為羚羊覺得獅子是“冷血的混蛋”；獅子也不會相信羚羊是“可憐的草包”。獅子想吃生肉，但當它們抓到獵物時，它們不會覺得“這太美味了”。這些詞匯——“混蛋”，“弱者”，“美味”——被哲學家稱作“厚重的價值詞匯”（Williams 1985），我們在第二章已經討論過諸如“德國佬”與“蕩婦”等輕蔑性詞語。[[47]](#_47_70)

“為什么我們不（認真地）使用‘蕩婦’、‘草包’、‘混蛋’這些詞語來描述動物的信念呢？”這是一個極好的問題，卻很少有人提及。為什么動物看起來沒有這些概念？

我對這個問題的思考，因為聽到電影《幽靈世界》的臺詞而變得清晰起來：

麗貝卡（對西摩說）：“伊尼德覺得你是一個呆瓜。”

西摩：“她是這么說我的？！”

西摩沒有問伊尼德是否喜歡或尊重他；看起來為了弄明白伊尼德是否覺得他是呆瓜，他需要搞清楚她是否用那個詞語來指他（或至少她是否有些愿意使用那個詞語）。如果麗貝卡對西摩的質詢這樣回應：“唔，伊尼德當然是不喜歡也不尊重你的。”那么，無論這令西摩多么沮喪，他大概至少能得到些許寬慰，因為伊尼德是否真的覺得他是個呆瓜還不一定。[[48]](#_48_68)

讓我們設想一個更直接的例子。比較一下教別人“德國人”與“德國佬”這兩個詞的過程。前者根本不需要提到語言。我們可以說：“某人是德國人，當且僅當他或她在德國出生或居住。”（要教其他純粹的描述性名詞，可以指向其涉及的典型對象，雖然光靠指向一個德國人不是很容易明確想要表達的特征。）與此不同的是，為了教會別人“德國佬”這個詞，可以用到教“德國人”用的全部說法，但是此外還必須加上一句，“‘德國佬’是一個冒犯性的詞”。——換句話說，必須要說起這個詞本身。實際上，在字典的明顯貶義詞條目里，這種分別體現得很明顯。《牛津英語字典》告訴我們“德國佬”有“貶義”——這并不是在描述這個詞所指的對象，而是在描述要通過“語義上溯”才能表達的詞語使用習慣。［“語義上溯”是一個W.V.奎因創造的短語，指的是從用詞語談論到談論這些詞語本身……例如從對英里數目的談論轉到對“英里”這詞本身的談論（1960：271）。］如果我們一一列舉出所有關于德國佬與德國人這兩個概念的常見描述，那么兩者的差別在于前者必然包括一些由語義上溯得出的描述；前者要談及“德國佬”這個詞語的使用習慣，例如“它是一個貶義詞；它是用來侮辱人的。”（關于“列舉常見描述”的討論，參見Smith 1994的第二章。）這個區分的重要意義在于，沒有語言的動物可能形成德國人的概念，但沒法形成德國佬的概念。當然，我認為動物其實這兩個概念都沒有，但它們沒有前者和后者的原因是不同的。動物不具有德國人這個概念，是因為要有這個概念，動物必須具有區分德國人與非德國人的能力（這可能包括確定他們在哪里出生，檢查他們的護照等——這些辨別能力動物根本沒有）。但是在原則上（在這個詞的某種寬泛意義上來說），沒有語言的動物可能具有這些能力。相反，為了形成德國佬這個概念，它們需要知道有關“德國佬”這個詞語的事情（比如它是用來表達輕蔑），因此，沒有語言的動物必然不能形成這個概念。對德國佬這個詞語的知識也許是隱含的、傾向性的、程序性的，或者是屬于怎樣做的知識而不屬于有關事實為何的知識；但對我們來說，最重要的是這是有關語言的知識。

另一個是來自弗雷格［（1897）1997：240—241］關于“狗”和“惡狗”（“cur”）區別的例子。達爾文可能是對的，狗確實具有狗的概念，但是沒有人會認為狗需要對“dog”這個詞（或德語的“chien”）有所了解才能有這個概念。科林·艾倫（Colin Allen 1999；同時參見Allen and Hauser 1991）論證道，一只動物要有概念X，它必須（1）能夠系統地區分X和非X；（2）能夠認識到自己對X與非X的分辨出了錯；（3）因為有（2）的能力，它還能學會更好區分X與非X的方法。也許狗確實有這些認識和區別狗的能力。然而我論證的是，要有能力形成惡狗這個貶義概念，任何個體（無論是人類還是狗）都必須對語言有所認識。如果我們假設（為了闡述更清楚，我們姑且把事情稍作簡化），“這是只惡狗”這個句子既表達“這是一只狗”的信念，同時也表達對這只狗的鄙視，那么我們就不能再說：要形成惡狗的概念，一個人必須能夠可靠和靈活地區分惡狗和非惡狗。我們沒法在全世界像挑出狗性一樣挑出惡狗性；因此，艾倫的標準可能在別的地方還算合適，但對于這種評價性的概念并不恰當。要獲得惡狗的概念，光是相信某個動物是狗而且對它表示輕蔑，這還不夠。畢竟，有一些狗似乎就可以滿足這個標準（我家的狗就很蔑視鄰居家的狗），但是我們仍然并不會嚴肅地認為它們具有惡狗的概念。要表達出圍繞著惡狗這個概念的陳規，就需要用到語義上溯。按照定義，知道這些陳規（雖然可能是怎樣做的知識）是掌握概念的必要條件，所以相關的語言知識對于擁有概念是必需的。

也許以上對于傻瓜、德國佬和惡狗的討論有些偏離主題，但這里的重點是，正如我在2.3節所論證的，有關這些貶義詞的情況同樣適用于大部分的道德詞匯。一個人要是不知道“德國佬”是個貶義詞，他就不能算是掌握了這個詞語，同樣地，一個人要是不知道他（認真地）使用的道德詞語表達了對某些實踐標準的認同，他就不能算是掌握了道德詞語。因此，一本完善的字典的道德術語條目，也應該能認可圍繞這些術語的、牢固的語言習慣。因此，圍繞道德詞語的陳規（知道這些陳規是掌握概念所必須）需要語義上溯才能得以表達。因此，道德判斷不能合理和嚴肅地被賦予非語言使用者。[[49]](#_49_68)所以，猩猩不會做道德判斷。

我們之前暫且承認了，猩猩的某些心理狀態可以合理地用其他類型的“應當”來描述。這同現在的結論沒有沖突。設想一下這些陳述句：

明天應當要下雨。但我并沒有接受任何認同明天要下雨的規范性標準。

你想要暖和點，而打開窗戶會令你變冷，因此，你應該關上窗戶。但我并沒有接受任何認同你關上窗戶的規范性標準。

這兩個陳述看上去都有點古怪，因為很難想象有誰會真這么說，但是這和第二章用數字標記的那幾個類似例子的古怪之處很不一樣。我們不會把“但是我并沒有接受”這部分視作抵消或撤回了前面的言語行為；這些例子的古怪之處與摩爾悖論沒有任何關系。因此，這些“應當”明顯不是表達譴責或贊揚的。我認為很多其他規范性詞語的情況也是類似，例如“這是最好的香蕉”或“這棵樹很好爬”里的比較性詞語。人們可以承認“托尼是附近最好的殺手”但同時不接受殺手行會的任何標準，同樣地，人們可以斷言“這是最好的香蕉”但同時并不表達任何意動性態度。因此，我論證的并不是我們永遠都不能使用規范性術語（比如“應當”“一定不能”“好”）來刻畫非語言使用者的信念，而只是在我們刻畫的信念里這些術語不能在道德意義上被特殊地使用。

## 3.3　道德信念的實在主義與工具主義

這個論證還需要進一步的簡要澄清。在日常用語中，“道德判斷”有時指一種心靈活動，有時則指語言的表達。這和我的論述沒有沖突，但是我已經論證過，正如我們之所以能認為別人持有“蕩婦”的信念，依賴于別人對“蕩婦”這個詞的理解能力，同樣，作為心智活動的道德判斷在某種意義上是從屬性的（這并不是說它因此就不太重要），也就是說，這種道德判斷的歸屬也取決于對方具有的相關語言能力。（可以說雖然我們傾向于把言談看作是出聲的思考，但是對于某些對象來說，情況剛好相反：思想是平靜的言說。）這里就引出了一個問題：我們是否能夠對于有這些內容的信念做實在主義的理解。

信念的實在主義者認為，當我們說喬相信p，我們的意思是說喬確實處于某種狀態（可以說是喬的大腦狀態），這種狀態的因果效力體現在他的行為中，并且（在原則上）我們能夠對其進行經驗研究。與此相反，工具主義者則認為，說喬相信p，只不過是表明賦予喬這個信念可以最恰當地解釋和預測他的行為（參見Dennett 1987[[50]](#_50_68)）。如果賦予某只類人猿心靈狀態可以最恰當地解釋和預測它的行為，那么對于工具主義者而言，這就已經等于說這只類人猿具有心智活動，但實在主義者頂多把這點看作是類人猿具有心靈狀態的良好證據。

我希望本書的讀者能理解為什么我想避免卷入這場復雜的、不必要的爭論。無論如何，我都認為這兩種觀點大致都是可以接受的。有些時候，我們把心靈狀態賦予一個個體，是站在實在主義的立場：我們說該個體確實跟某個可以稱為“信念”的東西處在一定的關系之中，而且這個信念同其他的心靈狀態，同自主的神經系統，乃至最終同世界都具有因果聯系。然而，我們沒有理由要堅持關于心靈狀態的言論總是深浸著實在主義的態度。如果在有些情況下我們想放松標準（說一個國際象棋電腦想要早點出它的王后，或一臺電腦相信它的打印機端口是空閑的），這也沒有什么不對。考慮和談論某些復雜系統的時候，把它們當作是具有心靈狀態可以給我們帶來便利，我們不必擔心它們是否真的具有這些狀態（關于實在主義與工具主義兼容性的一個詳細而有力的論述，參見Godfrey-Smith 2004）。

關于信念的一種常見和流行的實在主義觀點［杰瑞·福多（Jerry Fodor 1975，1994）與史蒂文·平克（Steven Pinker 1994）是其忠實擁躉］認為，信念乃是作為“頭腦中的句子”［即“心靈語言”（Mentalese）中的句子］而存在的。但是，心靈語言的詞匯怎么表達傻瓜、蕩婦或罪惡這些概念（能有這些概念的人需要熟知“傻瓜”、“蕩婦”或“罪惡”這些詞語在公共語言里怎么使用的）？說心靈語言的字典里有的條目是“貶義詞”，這到底是什么意思呢？我相信，答案是：心理語言并沒有這些英語詞匯的簡單對應詞。這不是說福多的理論無法解釋了，只是意味著，對于這些評價性的信念，他的理論需要能給出一套特殊的處理方式。但是，對于“露絲相信這只貓是棕色的”（實在主義者歷來關注的都是這種簡單例子）和“伊尼德相信西摩是一個十足的呆瓜”這兩句話，說它們的真值條件的性質完全不同，這應該也沒有什么出奇的。要形成“西摩是一個呆瓜”這個信念的條件（粗略地說）是有能力（或許還有意愿）把“呆瓜”這個詞肯定地用到西摩身上，如果我這樣想是正確的，那么以上的問題對實在主義就不構成障礙了。這是因為，我們可以合理地假設這種能力（以及這種意愿）是具有真正因果效力的狀態，而且對其可以進行經驗研究。

需要澄清的第二點則涉及工具主義的信念觀。如果把包含評價性詞匯的信念賦予非語言使用者，既有預測性也有解釋性的價值，那么我們這樣做有什么不可以的呢？比方說，賦予蜜蜂“黃蜂是十足的混蛋”這個信念，如果有助于我們的解釋工作，那么也許工具主義就會容許我們這樣做。為了回答這個問題，我需要對信念工具主義和信念實在主義之間的關系多說幾句。

盡管有時我們以工具主義的方式說話，比如說“這臺電腦認為打印機端口是免費的，但實際不是免費的”，但要說我們所有的心靈狀態歸屬都是這樣子的，或是說我們的言論從來不包含對實在主義的信奉，我認為都是極其不合理的（參見Lycan 1987對普遍化的工具主義的批評；同時參見Yu and Fuller 1986）。如果某人說，“這臺電腦認為打印機端口是空閑的，但其實不然”，而別人對此提出一個合適的嚴肅質疑：“那好吧，但是這臺電腦真的具有關于它打印機端口的信念嗎？”說話人總是可以收回或限定先前的說法。（這里的說話人當然不包括某些對具有信念或欲望要滿足的條件標準極其寬松的實在主義哲學家。）但是，涉及有關某個人的信念的斷言，就沒有這樣的收回傾向（當然，除非這個人是否真的那樣相信還是有尋常的疑點）。如果我宣稱愷撒相信盧比孔河里流的是水，有人對此提出質疑：“那好吧，不過他真的是這樣相信的嗎？”質疑者的問題只會把我弄糊涂；我不會明白提問者的用意何在。由此來看，工具主義看來不過是一種“方便的虛構”，我們有時候會容許自己參與這種虛構。比較一下更為人常見的虛構話語。假如我講一個童話故事，“公主住在一座由紅寶石做成的城堡里”，而有人要求我要以恰當的嚴肅方式對待自己說的話，那么我會非常樂意收回之前的說法。如果有人質詢道：“但是你不是真的認為有城堡是紅寶石做的，對吧？”我會回答：“當然不是了，但是你沒理解我的意思。”但是，如果我認真地說：“波希米亞公主伊麗莎白曾經與笛卡爾通信。”倘若這句話也遭到質疑，我就會堅持己見。我們清楚地知道，說“這臺電腦認為它的打印機端口是空閑的，但其實不然”跟說“薩利相信這臺打印機端口是空閑的，但其實不然”，兩者是不同的；如果第一句話受到質詢，我們一般都不會再堅持，承認那只不過是“修辭手法”。這些都表明，當我們在工具主義的基礎上賦予信念時，我們不過在進行無傷大雅且方便有用的虛構。

但是，如果以上對于工具主義在我們的信念賦予中的作用的理解是正確的，那么，擔心蜜蜂可以相信“黃蜂是十足的混蛋”就不是什么大問題了。如果我們賦予某只雄猩猩（他小心留意自己的配偶，因為后者對其他的雄猩猩來者不拒）“她是個蕩婦”這個信念（同時賦予這只雌性猩猩這樣一個信念，即其他雄性猩猩是猛男），這是沒問題的。同樣地，如果我們想要賦予猩猩以下這些信念：某些行為是被要求的，有些行為是應得懲罰回應的，有些性格特質是值得贊揚的；這也是可以的。從工具主義的角度這樣賦予信念不會帶來什么損害，只要我們清醒地意識到，這樣做只是一種偽裝，是一種為了交流方便而采用的捷徑。但是（正如我們假裝圣誕老人是從煙囪下來）我們不應該把這樣的說法太當真。至于蜜蜂是否真的相信黃蜂是混蛋，或者某只猩猩是否真的相信另一只猩猩是蕩婦或猛男，或者非語言使用者是否真的判斷某種行為是道德所要求的，等等，對于這些問題，答案統統都是“否”。

## 3.4　互惠的語言

在3.2節，我提出了一個先驗論證，結論是語言是進行道德判斷的必要條件。能不能論證相反的結論，即語言需要具備做道德判斷的能力嗎？我是看不出有什么希望可以用邏輯上的推理來這樣論證，但也許我們可以合理地提出理據，來證明事實上人類語言的發展可能正是為了讓我們進行評價。要了解這一點，我們必須轉向人類語言的演化。

對于我們的祖先開始用語言交流的準確時間，我們目前只能作理論推測。[[51]](#_51_68)需要記住，不能指望這個時間是非常準確具體的。究竟怎樣才算是“語言”，而不只是一套信號系統，不只是有意引起關注的聲音乃至恰好引起關注的聲音，這里的邊界是極其模糊的。即便我們神奇地獲得了有關人類祖先的全部觀察資料，包括所有神經科學的數據，能夠追溯百萬年來所有的發展過程，我們仍然會對語言何時演化出來這一問題有巨大分歧，因為這里的問題并非只是我們的知識有限（這當然是個問題），還是“語言”這個詞語的含義具有不可避免的模糊性。我們不應該被這種模糊性困擾，止步不前（我們的語言本來就充滿了有效而友好的模糊性），但是我們應該意識到它的存在。我們所能爭取的最好結果，是一套指出“在n年前，我們的祖先沒有形成語言，并且在n—m年前，他們形成了語言”的理論（如果有人問起一個中間的時段：“他們那時形成語言了嗎？”正確的答案是“多多少少吧”）。

比何時的問題，更有趣的是為何的問題。沒有任何理由認為答案必須是單一的；我們的祖先可能面對許多問題，而語言交流對這些問題提供了有用的解決方法。但這里我要強調的是關于語言最初“目的”的一個合理假說，這個假說主要來自人類學家羅賓·鄧巴與萊斯利·艾洛的研究（Aiello and Dunbar 1993；Dunbar 1993，1996）。[[52]](#_52_68)這個假說基于兩個經驗觀察。第一，我們可以注意到，人類以外的靈長類動物的社會凝聚力，依靠于受到嚴格調控的成員交互關系（常見的調控方式是相互清理毛發的行為），而個體只有固定時間分配給社會活動，否則它對其他重要目標的追求（例如覓食）就會遭受損失。事實上，這些靈長類動物個體最多只有20%的“社會預算”時間。第二，從觀察得知，在現存的社會性靈長類動物中，大腦新皮質的比例（即新皮質相對于大腦的其他部分的體積比）與群體規模之間有顯著關聯（群體規模越大，大腦新皮質占的比例就越高），這充分顯示大腦新皮質的主要作用之一就是處理社會信息。動物需要更大的新皮質不但是為了記住誰是誰，誰是誰的朋友或敵人，還是為了進行更準確和“高階”（“higher-order”）的心靈狀態賦予（例如，“她相信我知道他想要什么”），從而能夠更好地預測其他成員的反應（Mithen 1996：108）。

根據這第二項觀察，我們可以從對于人類不同祖先的頭蓋骨化石的研究，推斷出（臨時性的）他們的生活群體規模。通過這種方式，鄧巴與艾洛猜測，南方古猿（australopithecines）的群落成員數目大概有67人，能人（Homo habilis）有82人，直立人（Homo erectus）則有111人，以此類推。［至于群落規模增長的原因，則有不同的假說。例如理查德·亞歷山大推測，群落間的沖突對大群落更有利（1987：79—81）。］這些數據使得我們可以估計在每種人科動物需要花費多少時間為其他成員整理毛發，才能與聚落中群落里的每個成員維持交往：南方古猿所花的時間比重大概有18%，能人有23%，直立人有31%，尼安德特人（Homo Neanderthelensis）則有41%。結論是：在人類演化的過程中，相互往來的人們之間，有一股不斷增大的壓力，促進發展更新的和更有效的手段來交換有關彼此行為和關系的信息。這種壓力首先落在能人身上，隨后逐漸變得非常沉重，因此我們可以推測，人類的祖先最后肯定找到了某種解決方案。有人論證過，語言就是這種解決方案。要看一個人是否可靠，一個方法就是觀察她與別人的交往；但是，如果你想同時試探數十個潛在交往對象的可靠程度，那你就有麻煩了。但是，如果你同曾經跟他們交往過的人交談，你就可以很容易獲得所需的信息。花上20%的時間同其他人閑聊八卦，足以了解許多人的社會細節。同樣地，一個人也可以通過交談向更多的聽眾散布自我吹噓的信息（社交網絡早就存在了）。我們今天仍然耗費交流精力的一大部分進行“聲望管理”，這可以支持有關早期語言的“閑聊假說”。心理學家尼古拉斯·埃姆勒（Nicholas Emler 1990，1992）認為，這個比例高達60%—70%。杜魯門·卡波特為他的中篇小說《應驗的祈禱》（Answered Prayers，一部有嘲諷色彩的名人眾生相）辯護時，曾說過“我的書不是八卦，除非說文學作品全都是八卦”。我認為這句話包含了重要的真理。

此外，有了閑聊，一個人可以輕易地在遠處損害他人的名譽，這使得實施懲罰變得幾乎沒有風險，同時還極大地擴大了個人實施懲罰的影響范圍。（這意味著要解釋懲罰策略的演化，無須訴諸群體選擇的機制；參見1.7節。）這個關于語言演化的著名“閑聊假說”的內容等于是說，人類的語言能力之所以被選擇出來，是為了在人類群體規模增大的情況下方便互惠的交流。閑聊的語言就是互惠的語言。

在一個規模不斷擴大的大型群體中，互惠的重要性對語言的發展制造了日益增長的壓力，這不僅是因為閑聊的益處，也是因為語言使得更為復雜精細的交換活動成為可能，并且還促進了交換。勒達·克斯密德與約翰·托比（John Tooby 1989：64）對這一點做過精彩的論述：

如果我想用一個斧子交換別的東西，我如何表明自己想要什么呢？假設我指著你手中的梨子，以此示意我的目標。但我這樣指示意味著什么呢？我想要這只特定的梨子嗎？或任意一個梨子？五個蒲式耳的梨子？我想要某種水果，但不一定得是梨子。我想被帶到你可以發現這些好梨子的地方嗎？還是我想讓你扶住一架梯子，我好爬上梨樹？或者其他的果樹？還是說我想用自己的斧子切開你的梨，跟你交換半個梨子？

隨著物質文明的進步，交換具體的物品就變得越來越重要。這也可能間接地與群體規模相關，因為較大的群體容許和鼓勵勞動分工，而勞動分工可以促進技術的進步。那么，讓我們暫且采納以下的假設：最開始的言語交流同交換相關，這里的交換既包括安排復雜或明確的交易活動，也包括討論他人在交往中的表現（參見Paine 1967；Barkow 1992；Enquist and Leimar 1993）。

任何有關原初語言目的的假設都蘊涵了有關語言內容的結論。在閑聊和交易活動中使用的語言有什么基本構成要素？這種語言的成分應該能夠指稱具體的人，交易的物品和商品，乃至個體之間的關系。但我們也應該會預計到，評價性的語詞在一開始的時候就已經出現。我們的祖先不僅僅想要描述某人沒有報答他人的事實；他們聊天的目的是要譴責這樣的人。對于他人聲望的敏感性，從演化的角度看，其目的就在于避開、排斥、積極懲罰聲望極差的人。有效的（有料的）八卦閑聊不僅僅是為了描述誰對誰做了什么；它還包含了贊揚和譴責的語言——類似于“奧格一直沒有把斧子還給葛克：這個無賴！”（用個很老土的翻譯）或“克魯克總是以恩報恩：他真是個好人！”。從事廣告業的人們都會告訴你，評價性語言能更有效地自我推銷。

提起人類最初的語言，我們會自然地覺得，它應該主要由名詞和動詞組成（就像“我，泰山；你，簡”這樣的詞匯），而評價性的語言是后來才出現的。但是我的提議剛好相反：我們可以從鄧巴的研究推知，在人類語言剛出現的時候，評價性的語言已經是其中的核心部分了，原始人類語言的字典里就包括類似于“貶義”這樣的說法。這看上去或許像是過于空玄的思辨，但它直接來源于一個有充分經驗支持的原始語言功能理論。沒有人會認為，靈長類動物清理毛發的活動僅僅是為了除掉寄生蟲；這種活動是社會交往的媒介，是評估同伴以及評價自己和同伴之間關系的途徑。如果語言是群體規模擴大時用來替代清理毛發的工具，那么語言的功能應該也同清理毛發一樣。而且，既然我們能夠可靠地假定語言主要是通過其內容（而不是通過韻律或喉部的發聲）來實現這種功能，那么就可以由此推斷，語言在演化上的重要功能之一就是傳播某些類型的社會評價內容。

## 3.5　小結

斷定語言包含社會評價內容不等于說它包含道德語匯，因為我們在第二章已經看到了，道德判斷只是評價活動的一種特殊類型。這么想也許比較合適：道德詞語會作為討論名譽，安排交易，吹噓自己，譴責違規者等的工具而出現，這是因為，如果語言的目的是為了討論公正、欺騙、懲罰、應得、貪婪、所有權等，那要是缺乏“公正”“欺騙”“懲罰”“應得”“貪婪”“所有權”等詞匯（更準確地說，是可以被翻譯成這些詞匯的語詞），就很難有什么進展。然而，除非我們可以肯定他們給這些判斷加上了我們在2.4節談過的實踐力量（然而，現有的相關討論都無法保證這一點），否則我們還不能夠把人類祖先的咕噥發音翻譯成這些詞匯（從這些詞語的道德意義上來說）。我認為，事實上道德語言確實可以比非道德評價更有助于互惠交換（如果葛克說奧格違反了超越習俗的、無論喜歡與否都必須遵循的規范，說奧格應得懲罰，這比起僅僅是對奧格行為表示憤怒，能更有效地促使他的同伴們懲罰奧格的非互惠行為）。不過，對這個觀點還需要進一步的論證，我們留到下一章再開展。

讓我們在此稍作停留，總結一下我前面論述的幾個主要觀點。我們想知道人類使用道德判斷的傾向是如何以及為什么演化出來的。要處理這些問題，一個合理的做法是先觀察我們的靈長類近親（它們有些生活于復雜的社會系統中），看看它們具備和缺乏哪些道德的組成要素。我已經論證過，我們無需賦予猩猩任何道德判斷的能力，就可以解釋它們表現出來的遵循規則的行為。它們具有一定的抑制、嫌惡與行為傾向（無論是天生的還是習得的），對于某些行為（或某些不作為），它們會有負面的反應，而且如果不能在違規發生時就做出反應，它們會以后補上。它們還足夠聰明，可以預料到會遭受這種負面的反應，也會把心靈狀態賦予彼此。我也愿意承認，它們還很有可能“內化”了許多這樣的聯系，以致很多行為方式（可以算是親社會的行為方式）是出于習慣。德瓦爾給這種行為現象貼上“社會規則感”的標簽，看起來是合適的。為了論證的需要，我暫且承認了我們可以賦予猩猩含有假言“應當”的判斷，這樣一來，我就不覺得有任何理由要反對它們的社會生活在很大程度上的確是由“規范性”的規則考慮所掌控的。

但是道德在哪里呢？以上我們賦予猩猩的各種心靈活動，就算加起來，都不等于說猩猩認為負面的反應是應得的，或者把某個行為看作是違規的，或者判斷一個行為是恰當的，或者把某種性格看作是有德性的，把一次食物分配評價為公平的，或者相信一個物品是有所屬的，等等（即某些的個體對該物品享有特殊權利）。在本書的3.2節，我提出了一個論證來分析原因，來說明為什么事實上猩猩不能考慮這些東西。我的論證前提是：語言是形成這些想法的必要條件。我知道，有的讀者可能覺得我的分析沒有說服力，但即便如此，最主要的一般觀點仍然成立：猩猩可能具有形成道德的某些構成要素，但它們還缺乏其他的核心要素。我們應當關注這種演化的過程。人們是如何從抑制某些行為演化到對被禁止的行為做出判斷？是如何從不喜歡演化到不允許，從欲求某物演化到把某物當成是值得欲求的？當然，演化的過程是逐步的，但意識到這一點并沒有解決問題。一只紅色的動物可能是從黃色的祖先演化而來，然而，解釋了產生黃色的動力，再指出在黃色和紅色之間存在漸進的過程，我們仍然不知道從黃色到紅色的演變過程是如何發生的，以及為什么會發生。猩猩具有的道德構成要素（倘若完全從偏好、厭惡、欲望與抑制的角度理解）還無法拼湊或重組成道德判斷。（無論你怎樣組合黃磚塊都無法建成紅房子。）宣稱這種轉變發生在語言出現以后，說法雖然很有意思，但是并沒有回答問題。

直到目前為止，我都沒怎么提到情感的作用。我說猩猩的社會生活被抑制與偏好所控制，可能有人自然會覺得，這就等于是說猩猩的社會行為是被情感控制的。這種想法還伴隨著另一種常見的想法：猩猩之所以沒有道德判斷能力，只是因為道德判斷需要復雜的認知和概念能力，而情感無法滿足這些條件。我認為，這些為人熟知的想法總的來說是錯大于對，但我確實認為，情感是我們想要關注的演變過程的重要部分，因此我們現在就來討論情感。

許多情感都與我們的道德生活相關：憤怒、感激、義憤、反感、同情、鄙視、羞恥、負罪感、驕傲和幸災樂禍等——這個單子可以很長，但是我不打算一一討論，甚至也不打算闡述一套有關情感是什么的精確理論（參見Haidt 2003a；Fessler and Haley 2003）。我將要辯護的有關情感的看法，只需要細致到足以讓我表達我想表達的東西：以上列出的情感里至少有某些具有豐富的認知內容，而且評價性的概念對于形成某些情感而言是必需的。正因為如此，對語言是道德判斷的必要條件的論證之后，緊隨著的是，我們對情感的討論。從這兩個討論中可以得出一個有趣的結論：語言是某些情感的必要條件，而語言的演化使得某些道德情感成為可能。本章的余下部分將集中討論負罪感，但在此過程中也會提及其他情感。

## 3.6　情感

人們提出了許許多多的情感理論（彼此之間差別還不小，我們甚至會懷疑理論家們談論的是不是同一個東西）。羅伯特·普拉特切克（Robert Plutchik 1980）找到了多達27種情感的不同定義！值得慶幸的是，近些年來這個領域還是形成了一定的共識。情感（至少“基本”情感）被普遍認為是種適應機制，每一種情感都是由生物自然演化設計出來以完成相應的任務，調動生物體的生理、心理與行為等因素，從而促使它對生活環境里重復出現的、影響到繁殖適應性的威脅和機會做出適當的回應。按照這種觀點，大腦沒有所謂的“情感官”；每種基本的情感都“各行其道”，大體上不受其他情感的影響，它們（需要的話）會使用不同的神經結構，尋求自己特定的刺激，并且在特定范圍內產生行為反應。支持該觀點的證據來自有關情感的一些不言而喻的真理：眾所周知，情感的來去不受主體的自主控制；情感經常不管我們跟它講道理的努力；情感似乎自動地只對某些特定類型的刺激產生反應；情感影響我們的動機，繼而影響我們的行為。此外，在我們已知的歷史中，所有的人類文明社會所認識到的基本情感都是一致的——例如，古印度戲劇學家列出的“八種根本感覺”，就算是出自21世紀的心理學家之手也一點不出奇（Ghosh 1967：121—125）。[[53]](#_53_66)還有一些證據表明，有的情感反應可能因神經損傷而削弱，但同時傷者的其他情感還保持完好（Damasio 1994；Calder et al.2000；Adolphs et al.2003）。[[54]](#_54_66)除此之外，對于每種情感的功能，我們都可以建立可靠的演化假說，雖然理論家們還是就細節問題爭論不休。有的理論家從生理應激的角度來理解它們的優點（而我們能夠“體驗”情感只是因為我們恰好具有意識而產生的副作用）；有的理論家側重情感在記憶、動機、注意力、規劃和學習中的作用；有些則強調情感的交流作用。既然自然選擇機制偏愛一石二鳥的效果，我們完全可以設想，所有這些益處（以及其他的益處）都在情感的演化中占有一席之地，而不同的情感帶來的益處可能也有程度上的差別。

雖然對情感的達爾文式解釋極有優勢，但我們了解的每種情感不大可能都是產生于先天的、彼此獨立的機制（這種機制是為了解決人類祖先的某類問題而形成的）。更有可能的是，在演化形成的一套情感基礎上，文化的因素又參與進來，從而使人類的情感具有更新穎更精細的分類。這意味著在不同的文化和歷史時期里，人們會認識到出現不同類型的情感。例如，有一系列中文詞匯，能夠分別細微的情感差異，盡管它們被翻譯成同一個英文單詞“shame”（Wang and Fischer 1994）。但是，還沒有人類學家（雖然他們中有很多人都樂于接受這種挑戰）能提供例子來表明，其他文化認識到的某種情感對于我們來說是太過異類而無法理解的。我們能夠順利無礙地把中文里的相關情感詞匯歸為一類，稱之為“各種形式的shame”，這本身就很能說明問題。我們從來沒聽過有外文詞語是用來指某種“狂怒的同情”，這種同情里還略帶一絲嫉妒，是由下雪引起的，又能令人發笑。情感之所以具有跨文化的多樣性，是因為在不同文化里，不同情感的重要性有所不同（例如羞愧這種情感有時是核心組織原則），文化使得情感的誘因有差異，也使得情感導致的行為有差異，不同文化對情感領域的分類方式不一樣，從而使得其中的情感多少有些比較細致的區別。但這些觀察事實都不能推翻以下假說：在這些差異背后，存在著一組先天的普遍情感。有不少論證都是關于到底哪些情感先天的或“基本的”，關于人類的情感生活有多大部分是在先天限定的基礎上由文化因素增添的（當然還是有些人抱著落伍的觀點不放，認為情感生活全都是文化性的）。對于這個經驗問題，答案依然懸而未決，但本書表述的論證并不要求我們在這場寬泛的論辯中有所偏向。另一個需要注意的重要之處是，所有這些演化論上的思考并沒有解決情感到底是什么的問題（可以稱之為情感的本體論問題）。情感是感覺，是神經事件，是行動傾向，是信念，是跟以上這些完全不同的什么東西，還是它們的某種綜合？對于這個問題，我們也沒有必要選擇任何立場。

我希望著重強調的是，某些情感具有一定程度的認知復雜度。考慮一下厭惡感。心理學家保羅·羅津（Paul Rozin）和他的同事曾經論證道，“核心的厭惡”與拒絕進食有關，這也是達爾文的觀點［（1872）1999：255—260］。臉部特定的厭惡表情像是要把食物從口里吐出來，在極端情況下，厭惡可以引起作嘔，哪怕并沒有吞食厭惡的東西。然而，羅津等人（2000）論證說，厭惡并不僅是反感：厭惡包含有被冒犯與被玷污的感覺。后者被不少實驗證實：被實驗者不樂意穿陌生人穿過的毛衣，即便那件毛衣已經被細心洗過了。如果他們相信衣服之前的主人有一條腿被截肢了，或者是殺人犯，他們穿這件衣服的意愿會進一步急速下降（Rozin et al.1994）。羅津（1999：33）寫道：

厭惡和交感式迷信思維（sympathetic magical thinking）的聯系非常緊密。令人厭惡的物品會污染它們接觸過的東西，這體現了交感迷信的法則：“一旦有接觸，永不能祛除。”［Frazer（1890）1925：11］……人們對所厭惡之物的反應如此強烈，以至于厭惡常常波及到本身并非令人厭惡，只是形似被厭惡之物的東西。因此，人們對于做成糞便狀的塑料或巧克力……或人工合成的鼻涕狀黏液，也會有厭惡感，即使他們知道它們并不真是看上去那樣。

在此基礎上，羅津等人（2000：646—647）得出結論：動物與嬰兒沒有厭惡感，因為“能夠形成看不見的物體和表象不同于真實這些觀念，是相當抽象的認知成就……認知能力上的限制是阻礙兒童完全具備厭惡感的主要原因”（同時參見Rozin and Fallon 1987）。有觀察證據支持這個假說：雖然動物和嬰兒顯然會反感某些食物，但他們不會表現出那種與厭惡相聯系的、由被冒犯而產生的極度排斥和憎惡感。

負罪感（guilt）是另一種兒童較晚才具有的情感。這完全不會推翻負罪感是一種先天情感的假說，因為并不是所有“設計好的特征”都得是生下來就有的（比如青春期的發育）。快樂、悲傷、憤怒和恐懼似乎是最早出現的情感（Emde 1980），當兒童開始使用語言時，相比其他情感，這些情感的名稱也是他們更早理解和使用的。一項研究發現，超過90%的兒童在他們30—35個月大的時候就理解了“害怕”這個詞，但要遲至54—59個月的時候，只有23%的兒童理解“guilt”這個詞（Ridgeway et al.1985）。當然，在兒童具有會使用“guilt”這個詞語之前，他們可能已經有過負罪感的經歷，而且年幼的嬰兒可能具備了這種情感的一些基本要素（參見Zahn-Waxler and Kochanska 1989；Barrett 1995）。但是，我們還是可以合理地認為，負罪感出現在其他的情感之后，而這很可能是因為負罪感就像厭惡一樣，包含一定的概念能力，這些能力需要經過一段時間才能形成（參見Kagan 1984；Damon 1988；Lewis 1992）。盡管負罪感同很多其他的復雜性狀一樣是逐漸發展起來的，這個結論還是可以成立的。在不太嚴格的對話中，我們當然可以指著一個擁有部分負罪感要素的兩歲孩子，說她有“負罪感”，正如我們可以指著一個能說“mama”“dada”的一歲孩子，說她是在“說話”。但是，在后一種情況下，說一歲的兒童滿足了語言能力的充分和必要條件，說所有她今后發展起來的，組合復雜詞語的能力嚴格來講都不是語言的必要成分，那是非常瘋狂的。同樣，假定認知要素對于負罪感和厭惡而言是不必要的，也是錯誤的。我們的情感詞匯主要還是用來形容正常的成年人，雖然我們偶爾會為了談話的方便，把這些詞匯應用到其他達到這些情感的部分標準的主體上，但這不等于說它們還沒有達到的標準根本就算不上標準。

我指出某些情感需要概念上的復雜性，并不是要支持那種認為這些情感需要信念為條件的觀點。許多人認為，信念是情感的核心部分。事實上，古代斯多葛學派認為，情感就是信念。我認為這個觀點是錯誤的，如果你不同意的話，那就去看一部不錯的恐怖片。（對古代斯多葛學派也是類似的建議：去看一出不錯的悲劇。）假如這是部吸血鬼電影，還讓你覺得害怕。有些哲學家甚至會干脆否認這種可能性。他們牢牢認定情感必須包含信念，如果被要求考慮觀看虛構作品（沒人把看到的內容當真）的情況，他們就徹底否認我們在觀看時具有真正的情感：恐怖電影并不令我們恐懼，悲慘的小說并不令我們悲傷，惡俗喜劇并不令我們感到惡心。接下來他們就會告訴我們，虛構人物不可能是性感或有趣的！常識當然可能出錯，但倘若用來證明常識為錯誤的只是一個不牢固的哲學論點（它完全不解釋為什么常識被嚴重誤導），那么最有可能的是哲學家自己搞錯了（關于對這一點的詳細論述，參見Joyce 2000a）。

但是我們對虛構作品的情感反應仍然令人感到困惑。在恐怖片的例子里，令我們感到恐懼的對象到底是什么？你可能會說，吸血鬼的形象讓你聯想到真實的威脅，例如連環殺手，而后者才是你害怕的真正對象。但這個觀點似乎是錯的。你當然不曾意識到你是在聯想連環殺手（你害怕的只是嗜血兇殘的吸血鬼），因此為支持這種“記憶聯想理論”的人必須認為，你害怕的真正對象存在于你的潛意識。這種觀點并沒有不一致，但是很不合理。不，你害怕的就是吸血鬼，雖然你并不相信他們存在；你片刻也不相信有任何東西（或任何真實的人）在威脅你。你要是真相信吸血鬼存在，你就會給自己找來一串大蒜和一個木釘。然而，即便你并不相信吸血鬼要吸你的血，你仍然有這個想法。你在床上輾轉反側，緊張地聽著樹枝拍打窗戶的聲音，你是在想著但并沒有肯定“吸血鬼追我來了”這個命題。我們可以說：你是因這個想法而感到害怕，但是你所害怕的并不是這個想法本身（Lamarque 1981，1991；Joyce 2000a）。作為具有內容的心靈狀態（即作為意向性狀態），這個想法提供了你害怕的內容：吸血鬼。作為大腦事件，這個想法導致了你的恐懼。需要注意的要點是，鑒于以上論證，想著p和相信p都使用了同樣的概念。因此，虛構作品能引起情感上的反應，這并沒有給“有的情感要以復雜認知為前提”這個主張造成困難。

人們能夠對虛構作品產生任何一種情感嗎？這似乎是一個我們不必關切的問題，但事實上，這個問題的答案對于本書末尾將要討論的問題非常重要，現在正是處理它的好時機。表面看起來，有一些情感確實是虛構作品無法引發的。嫉妒常被認為是其中之一（Radford 1995；Neill 1995）。但事實上，我看不出一個人為什么不會對虛構作品有嫉妒的反應，雖然我承認這不常見。假設有一個單身漢想象他有個女朋友，然后繼續（有點沮喪地）想象她的愛被另一個更年輕有活力的追求者奪走。這樣的想法難道不會令他感到生氣嗎？他感受到的為什么不能就是嫉妒呢？在我看來，對這兩個問題做肯定答復是最自然的。這個沉湎于想象的人仍是在嫉妒某個人，但他嫉妒的不是真實世界中的人；他并不相信確實有個人正在奪走自己的心上人，但是他可以有這個想法。嫉妒虛構人物是不太尋常的現象，因為嫉妒是不能間接體驗的，這種情感要求設想自己和別人之間有關系（一種特定的關系），但書或電影通常都不需要激發這種設想。閱讀《安娜·卡列尼娜》可能需要讀者假裝安娜存在，但是通常都不需要讀者假裝自己和安娜有互動。我不會花時間考慮跟安娜聊些什么，或者她會怎么看待我的衣著，或者我能否在19世紀圣彼得堡喝上一杯上好的意式濃咖啡。這并不是否認有人可能會這樣設想。有人也許太喜歡這個角色，開始幻想自己與她展開一段關系，于是當渥倫斯基出場時就心懷醋意。我的意思只是說，虛構作品一般都不會引導讀者往這方面想象，所以這樣的反應是很不尋常的。

因為觀看虛構作品而感到負罪感，同樣也是罕有的事情。因為負罪感的產生通常是由于相信自己對別人做了錯事，然而虛構作品一般不會讓人進行這樣的假想。但是正如前文所述，負罪感似乎并不必然包含信念。如果我坐下來設想我背叛（不論是想象的還是真實的）朋友的可恥行為（不一定是我完全不想做的行為）的種種生動細節，那么我似乎可以產生些許的負罪感之情。這種負罪感不同于在真實生活里犯錯的負罪感，就像對好萊塢吸血鬼的恐懼不同于對死于車禍的恐懼，但是說前者“不同于”后者并不等于說“前者是完全不同類型情感”或“前者根本算不上真正的情感”。日常語言使用同樣的詞匯描述這兩種情形，我們對情感的神經機制所掌握的經驗證據都告訴我們，常識在這個問題上是正確的。如果情感是生物自然選擇的產物，目的是應對現實生活里的問題，那么我們可能會疑惑為什么光憑想象就可以引發情感反應。一種簡單的解釋，用兩個杰出的心理學家的話來說，是“促發情感的認知評價活動相當粗略，不做任何真實性的確認”（Johnson-Laird and Oatley 2000：465）——我覺得，這就好比人類引發性欲的系統不需要確認引發興奮的是真人還是雜志里的彩圖。除此之外，另一種解釋則認為在祖先環境里，享受虛構活動的傾向對人類的適應性有所裨益。[[55]](#_55_66)

關于情感，我想再做最后一個一般性的評論。我想要再次強調，上文的內容在情感是什么這個問題上并沒有采取任何立場。唯一被排除的理論類型是那些主張信念是情感的必要條件的立場——因此，情感不可能是信念。（斯多葛學派的心理學就到此為止。）我的主張只不過是說，對于某些情感，使用概念——有時在信念里，有時是在思想里——是必須的。至于“怎么個必須法”，我不會給出任何答案，因為本書持何種觀點并不依賴于如何回答這個問題。但是有一點值得指出：某類情感必然包含特定的信念或想法，從這個前提并不能得出結論說我們在（用“憐憫”、“厭惡”與“負罪感”這些詞）指稱情感時，一定程度上也是在指稱那些信念或想法。為了說明我的意思，可以考慮一下傷疤的概念。在我右手掌心上，碰巧有三個看起來很像傷疤的印記。但它們不是真的傷疤，因為是我生下來就有的；它們不是由外界導致的損傷。能算成傷疤，需要有一定的成因。這是先驗必然的；這是“傷疤”這個詞的意思。盡管如此，造成傷疤的事件并不屬于“傷疤”這個詞指稱的對象（事實上，很難理解這個事件怎么可能是“傷疤”指稱的對象，因為它是造成傷疤的原因）。當我指稱“我腳踝上的傷疤”（那里確實有個傷疤，是我抬腳穿過窗戶時留下的），我指的我腳踝上的東西（可見可觸摸的東西），而不是某種由一個物體和多年以前打碎玻璃的事件組成的奇怪混合物（參見Thalberg 1978；Shaffer 1983；Green 1992）。

信念/思想與情感之間的關系或許就類似于使身體受傷的事件與傷疤之間的關系；信念/思想是情感詞匯指稱對象的必要伴隨物（和原因）。如果這是正確的，那么我們幾乎就沒怎么涉及情感實際上是什么的問題。情感可能是一種感覺，也可能是某種神經或生理事件。另一方面，也許信念或想法是情感的構成部分（即情感是種混合物）。也許信念或想法就是情感本身，而相關的感覺或生理事件不過是情感不必要的伴隨物。無論這個問題有多么引人入勝，在這里尋求答案只會讓我們偏離主題。

## 3.7　負罪感

說到負罪感，有意思的一點是它和羞恥、尷尬一樣，都屬于自我導向的情感，都是“從內部”引導人的道德行為。我們前面大部分的討論都是集中在他人導向的道德判斷（例如葛克譴責奧格偷他的長矛），但自我約束顯然是我們道德生活的重要特征，因此也需要解釋。事實上我們甚至會覺得，要是一種生物沒有把規定“內化”，說它具有道德感就是不合理的；而且，除非它有把這些規定用于自身的傾向，否則我們不能說把規定內化了；也就是說，除非它具有良心。達爾文認為，把“道德感”等同于良心是合理的：

……任何具有明確社會本能（包括父母子女以及家庭成員間的親情）的動物，一旦其智能發展得像人類一樣完備（或接近人類），就不可避免具有一種道德感或良心。”［（1879）2004：120—121］

這段話的上下文的語境表明，“道德感或良心”這個析取顯然不是用來指兩個不同的東西，而是在表達大致的同義詞或指稱相同的術語；就好像說“紐約或大蘋果”[[56]](#_56_66)，而不是“紐約或費城”。在此之前，達爾文還曾大膽地斷言，“在人和低級動物的所有差異中，道德感或良心是最為重要的”（120）。

達爾文的第一個說法給我們出了個難題。親社會情感加上“智能”，怎么就產生了道德判斷？為什么“智能化”相當于“道德化”？達爾文有一點是非常正確的（我在2.1和2.2節也強調過這一點）：親社會的傾向還要加上某種東西，才能真正談得上是“道德感”，因為產生良心需要有相當的認知復雜度。雖然我們也說良心的“刺痛”或“傷痛”，但這只是比喻而已；負罪感并不是真的像是被針扎的感覺。具有良心不僅僅是不想做或不喜歡做某些事情，有良心的人不想那樣做是出于一定的理由：是因為判斷那些事情是違規的。“負罪感”既指一種情感，也指人和違背規范之間的關系（比如“他對罪行負有責任”）。因為當一個人有負罪感時，他也判斷自己對錯誤負有責任。沒有這樣的判斷，欲望或抑制無論有多么堅固牢靠，都還無法構成負罪感這種情感。

我們至少已經可以處理達爾文難題的一些要素。他說智能化相當于道德化，這看起來有些奇怪，因為我們會很自然地覺得，良心大多數時候是通過負罪感體現的，但負罪感是情感，不是“智能”。然而，如果我前文的論述是正確的，這種自然的想法其實來自一個不恰當的二分法。負罪感當然是一種情感（這我沒有否認），但是這種情感包含“智能”因素，“智能”在這里指運用概念的能力。

負罪感所負載的概念內容是什么？先比較一下負罪感與羞恥，或許會有點用處。很自然地同負罪感聯系在一起的，是判斷自己犯了錯（這種錯誤可以得到彌補）；同羞恥聯系在一起的，則是判斷“自己有問題”，而且也許對此不能改變什么。一個人因為說謊而有負罪感，因為自己是個騙子而感到羞恥。負罪感涉及的是違規行為；羞恥感則來自缺餡。負罪感會讓人彌補過錯；羞恥感則會讓人想要退出社會活動。人們可能會因自己出身卑賤、鼻子太大、偶爾輸給不如自己的對手而感到羞恥，但不會因這些事情而有負罪感（有幫助的討論以及進一步的參考資料，參見Tangney and Fischer 1995；同時參見Niedenthal et al.，1994）。雖然從直觀上來說，這些例子對區分負罪感和羞恥是很有用的方法，但要截然分開這兩者卻是沒有什么意義的，況且在很多情況下，我們可以恰當地說某個人同時具有這兩種情感。有些語言似乎沒有對這兩種情感做什么區分（Levy 1973：42），在有些文化里負罪感不如羞恥感復雜。但這兩個觀察事實都不意味著這些文化沒有負罪感。人類學家覺得區分“恥感文化”（“shame cultures”）與“罪感文化”（“guilt cultures”）有時候很有用，而且這樣做偶爾還預設了前一種文化（其中羞恥感極為重要且詳盡繁復）根本不承認負罪感。但是，當我們回頭看看這個區分的理論起源，即人類學家魯思·本尼迪克特的著作，就會發現她是把羞恥感同“內化了的對自己有罪（sin）的確信”對立起來的（Benedict 1946：223）。跟罪與救贖的框架聯系在一起，這只是負罪感在文化上得到詳盡表達的一種方式（這也是基督教傳統的重要部分），但要是假設負罪感就只能具有這種形式，并由此斷定在不具備那種框架的文化都沒有負罪感概念，就太過狹隘了。

心理學家刻畫的負罪感要更為復雜（參見Zahn-Waxler and Kochanska 1988；Baumeister et al.,1994）。他們揭示了導致負罪感的最重要誘因是對他人的傷害（Hoffman 1982；Tangney 1992），一個人負罪感的強烈程度，同在他看來自己與受害者的社會關系有多重要成正比（Fiske 1991；Baumeister et al.1994）。因此，負罪感通常與違反道德規范的行為聯系在一起，就一點也不出奇（Tangney 1992）。我們都知道，人們也會對非道德的和自我導向的違規行為懷有負罪感，但是這類負罪感需要有很深的社會化程度才能逐漸形成，而且最容易解除，這些事實都表明它是“核心負罪感”最末端的衍生產物。

弗洛伊德（1929：123）宣稱，負罪感“表達為對懲罰的需要”，這是個敏銳的觀察，雖然他也對負罪感的發展過程說過許多蠢話。需要注意的是，這不同于對懲罰的恐懼。人們可能對懲罰感到恐懼，但不會帶有負罪感；也可能有負罪感，但不害怕懲罰。一個人也許知道她違反了某些規則，害怕將會遭受懲罰，但是除非她在某種意義上認同這些規則（即便她從整體而言并不接受這些規則），否則她不會產生絲毫的負罪感。要有負罪感，她得要判斷自己應得懲罰，而不僅僅是將會遭受懲罰。我認為，這比弗洛伊德說的“負罪感是對懲罰的需要”更為恰當，因為人們有時會感到懲罰是必需的，但不是應得的；或者判斷懲罰是應得的，但不覺得需要實施懲罰。

上一段括號里的內容值得注意：一個人可能接受某些規則，但同時并不是從整體而言接受這些規則。一個前羅馬天主教徒可能對避孕懷有負罪感，盡管她已經不再支持天主教的信條。但很明顯的是，這些價值觀念還是對她很有影響；“避孕是一種違規行為”這個想法仍舊活躍在她的心里；她心靈的某些角落依然接受這條規定，即便它并不代表她全面考慮的整體判斷。把負罪感看作不過是某種不含概念的焦慮，或是環境造成的純粹厭惡，這種看法是經不住仔細推敲的。當然，有的前天主教徒可能只剩下對某些活動的非認知性厭惡，但如果她沒有違反規范，或應得懲罰，或應得責備的念頭（哪怕只是從她心靈深處某個角落里傳出的惱人低語，違背她更好的判斷），她的感受就不能算是負罪感。類似地，對于負罪感包含“違反規范”的想法，或許有大量資料記錄的“幸存者負罪感”可以作為反例，但是我同樣會強調，雖然這些人站在理性角度，非常清楚他們對別人遭受的傷害沒有責任，他們還是不能完全擺脫他們“做了錯事”的感覺。正是因為他們的感受具有這種現象特征，我們才傾向于把他們的痛苦稱為“負罪感”。

從前文的討論，我們可以得出兩個結論。首先，如果負罪感必然包含自我導向的道德判斷，那么只要我們認定負罪感是一種先天機制，我們就得推斷說，進行道德判斷的能力也是先天的。但是這個結論也許太強了，因為我們要意識到，違規行為還包括對非道德規范的違反。更為保險和謹慎的結論是：只要我們認定負罪感是一種先天的機制，那么我們就由此推斷說，人類具有先天的能力，來判斷某些行為違反了得到認同的規范框架，因而應該受到修正或懲罰。其次，我們現在可以把本章前后兩部分的成果綜合起來，得出結論：非語言使用者不可能有負罪感。語言是賦予某些評價性概念的前提，而這些概念對于某些道德情感而言是必需的。我小時候養的狗（縱然它有服從的行為，也知道潛在負面后果）沒有也不可能有負罪感。也許在它那犬科動物的大腦里，有一些與人類負罪感相似的“某種東西”，這“某種東西”可以影響它的行為，甚至可以算得上是情感。但是它不是負罪感。對我們而言，更好的做法是把它稱作“前負罪感”（proto-guilt），或把它當成一種新的情感類型：“狗負罪感”或“動物負罪感”。這樣做可能太過學究氣，以至于顯得有點傻氣，但我并非是在論證日常對話也要這么精確。我論證的是，當我們以仔細審慎的方式談論動物的認知能力時（這正是我們在本書中做的事情），需要限定乃至放棄這些詞匯。

按照一個常見的簡單假說，人類祖先的社會生活是由一些親社會性的情感調控的，比如愛、憤怒、鄙視、厭惡和負罪感，然后不知怎的，運用復雜抽象概念的能力就降臨了，這使得道德判斷成為可能。現在可以看到，這個假說很成問題。事實上，某些親社會情感的出現，某些概念的“降臨”，以及語言的進展，這些因素是結合在一起的，因此導致人類道德感的演化也頗為錯綜復雜。這并不是否認某些親社會的情感（例如愛與同情）先于語言與道德判斷出現（我基本上在2.1節論證過這一點），因為我沒有論證一個很強的觀點，即所有情感都要以概念為條件，因此我也沒有論證所有情感都要以評價性概念為前提條件。我很懷疑能否對我們歸入情感的所有事物給出一套整齊統一的分析（盡管我承認過去曾試圖這樣做，參見Joyce 2000a）。我有把握指出的只不過是：作為道德良心的核心情感，負罪感伴隨著評價性概念。如果要被追問還有什么其他的情感也是如此，我首先想列出的會是羞恥、義憤以及“道德厭惡”。

但是，正如前文所提到的，這些情感的出現同語言和道德感的演化密不可分，這個事實仍然沒有解決一些重要問題。我們不能就簡單地設定，在語言降臨的時候，“應得修正或懲罰的違規行為”這種想法也出現了，好像背后有魔力控制一樣。我們仍然需要了解產生那種想法的傾向為什么會形成，以及是如何形成的。這樣，也只有這樣，我們才算有了一套關于道德判斷的能力是怎樣通過生物自然選擇發展出來的假說。我們現在就轉向這些問題。

# 第四章　道德感

## 4.1　預覽

道德思維有什么用途？無論我們對道德感的演化看法如何，這個問題都值得研究。明確這個問題的時態很重要；我們想了解的不是道德現在如何有適應性，而是它如何成為了適應性狀——換言之，我們想知道它過去如何有適應性。要想心安理得地宣布某個性狀是自然選擇的產物，就必須就該性狀為何對我們的祖先有用（即提高其繁殖適應度）這一問題提出一個合理的工作假說。（當然，光有工作假說還不夠；就算把自然選擇塑造某一性狀的過程說得合情合理，這也不能證明該性狀就是那樣產生的。）至于道德是否依然有用，則是另一回事。某些極端的思想家持否定態度。（尼采或可作此理解；另見Hinckfuss 1987 and Garner 1994。）也許進行道德判斷的傾向就類似于對甜食的喜好：在古代，甜食少見并且通常富含營養，因此對甜食的喜愛是適應性狀；但是在甜食泛濫且易得的時候，這種喜愛對很多人就成了致命的威脅。這種可能性不是沒有，但我認為微乎其微。理解了道德思維如何提高我們祖先的繁殖適應度，我們就很可能會認識到它是如何仍然對我們有用的，反之亦然。（這不等于說我們就會認識到道德依然是有適應性的；上句中的“有用”是個日常用語，與基因的復制能力無關。）

為明晰起見，我將這個問題分解成幾個子問題（不過后面會闡述合并其中兩個子問題的理由）。首先，道德感是通過群體選擇還是通過個體選擇演化的，這是個問題。如果群體選擇是主要過程，我們必須問為什么個體的道德思維可以有益于群體；如果個體選擇是主要過程，我們則要問為什么道德判斷可以提高個體的適應度。其次，我們要區別對他人做出道德判斷所帶來的益處和用道德慎思調控自身行為所帶來的益處，因為這兩者完全有可能是為了不同演化目的而產生的適應性狀。因此一共有四個問題值得研究：

對他人做出道德判斷如何有益于判斷者所屬的群體？

對他人做出道德判斷如何有益于判斷者自己？

對自己做出道德判斷如何有益于判斷者所屬的群體？

對自己做出道德判斷如何有益于判斷者自己？

需要注意的是，道德先天論者（moral nativist）沒有必要對四個問題全部做肯定回答；一個肯定的回答就足以捍衛他的觀點。但其實我認為四個問題的答案都是肯定的。第四個問題直觀上最有挑戰性，所以接下來（4.2節和4.3節）我將主要概述對這個問題的解答。我們在此感興趣的問題是“為何”：為何自然選擇會演化出道德感？（為何道德感可能是適應性的？）本章的第二個任務是把話題引向與“為何”自然相關的“如何”：道德感是如何演化的？——自然選擇讓人腦發生了什么變化，使它具備了道德判斷的能力？4.4節將提出如下假說：道德感的核心是人類將情緒投射到世界上的傾向。至于更進一步的問題，“自然選擇讓人腦發生了什么變化，使它具有投射情緒的能力”，這個問題雖然合理，但我還無法解答。隨著我們接近關于大腦的現有知識的極限，解釋工作最終也就只能到此為止了。

完成以上兩個任務以后，我們將得到一個關于道德演化的“原來如此”的故事。但我不滿于就此結束，因此，本章的第三個也是最后一個任務，就是綜述一些支持這個故事的經驗證據，從而讓這故事成為在科學上可以有一席之地的假說。

## 4.2　作為個人承諾的良心

對自己做道德判斷，也就是所謂有良心，為什么能提高判斷者（相對于不具備這個性狀的競爭者而言）的繁殖適應度？我對這個問題的看法主要來自以下自然的假設：如果某人對于一個可行的行為作出了正面道德評價，那么他做出該行為的概率就會提高（反之則降低）。如果某行為（或不作為）能提高繁殖適應度，那么任何確保該行為（或不作為）得以實施，或提高該行為實施概率的心理機制，也（相較于更難以達到這些結果的機制）更能提高繁殖適應度。因此，自我導向的道德判斷只要與恰當的行為形成關聯，就可能提高繁殖適應度。如前所述，在許多場合“恰當的行為”（即提高適應度的行為）包括助人與合作行為。所以針對親社會行為——針對判斷者自己的親社會行為——的道德判斷，有可能提高判斷者自身的適應度。

我們的關注點應該放在“道德判斷影響動機從而提高行為的概率”這個前提上。當代許多哲學家持所謂“動機內在論”（與動機外在論相對）的觀點：道德判斷先驗必然地包含行為動機。人們通常把內在論和非認知主義聯系在一起，不過我在別處（2002；forthcoming a）論證了這種聯系其實是錯誤的。我在第二章曾經指出，非認知主義者主張道德判斷表達某種意動的、涉及動機的狀態，但此處的表達關系不是因果關系，而是約定俗成的關系。因此，正如道歉表達悔意并不意味著所有道歉都伴隨著真誠的悔意，非認知主義并不會導致“道德判斷必然伴隨動機”的結論。不過，我在2.3節還是對非認知主義表示了有保留的支持，而我支持的東西確實蘊涵“道德判斷和動機之間存在某種先驗關系”這個結論：“標準的道德判斷”（“canonical moral judgement”，不妨如此稱呼）的確伴隨著動機狀態。在一個標準的言語行為中，行為者在說話時要處在該言語行為表達的所有狀態。因此，如果公開的道德判斷“盜竊有錯”，既表達盜竊有錯這一信念，也表達對譴責偷竊行為的某種標準的意動接受，那么該判斷的標準個例就需要判斷者兼具這兩種心靈狀態。因為“意動接受”是一種涉及動機的狀態，所以標準的道德判斷都包含動機。[[57]](#_57_64)

但是即便你不接受以上元倫理學觀點，你也可以同意道德判斷影響動機——只要你承認道德判斷和動機間存在完全偶然的心理關聯。至于這種關聯的存在，從意志薄弱現象可見一斑。雖然人類擅于計算長遠利益，但眾所周知，我們在使自己的動機與慎思的結論相符這方面做得極為糟糕。就算明知一時快活的代價是長遠損失，我們還是抵擋不了誘惑。我們吃各種垃圾食品，熬夜，不到最后一分鐘不改學生的論文，買了健身器材卻閑置不用，凡事能拖則拖。雖然各人糟糕程度不同，并且無疑也有少數神人可以理直氣壯地宣稱自己對此類惡習免疫，但不能否認，不能讓動機與審慎的判斷相符，放任一時的誘惑淹沒內心實踐理性的聲音，是人類心理的一大特點。

同時我也得承認，為了強調一種現象，上面只說了事情的一個方面；另一方面，人與其他動物比起來當然審慎得多。但在我們審慎思考的結論與動機和行動的形成之間無疑存在一個重要缺口。支持某一行動的審慎考慮（“為了健康，我應該點沙拉”）就是無法可靠地轉化為實際行動。我們多數人對此有親身體驗，相關的經驗科學證據也是汗牛充棟（Schelling 1980；Elster 1984；Ainslie 1992）。這需要解釋；如此頻繁地妨礙我們追求自身最佳利益的設計缺陷，為什么沒有被自然選擇糾正？大概是因為不同心理官能各有各的計劃，從而在行動上相互掣肘。但盡管“官能競爭”的觀點很可能是正確的（連柏拉圖都這么認為），自然選擇為什么沒有糾正這個嚴重的缺陷，這仍然是個問題。比較可能的答案是，意志的薄弱是為了實現其他有價值的目的而必須付出的代價。那個目的或許就是靈活計算主觀偏好的能力。能夠隨情況變化重新評估香蕉價值的生物通常比與永遠視香蕉為最高價值的生物更有優勢。飽的時候一根香蕉對你沒有什么價值，但餓的時候它也許是你能得到的最有價值的東西。但犯錯是任何這類靈活的全面評價系統［即實踐智力（practical intelligence）］不可避免的代價。正如自然選擇造不出既擁有靈活的信念生成系統又永遠只產生正確信念的生物，它也造不出既有靈活評估事物主觀價值的能力又永遠只作出正確評估的生物。按照這個觀點，意志薄弱是這類實踐錯誤的個例：眼前利益的誘惑（這誘惑本身就是某個先天機制的產物）導致個體錯誤地把一個壞選擇給“合理化”了。

讓我們設想有一類行為總是至關重要的，以至于自然不想把這類行為的成敗托付給普通人脆弱的實踐智力。這類行為也許是（比如說）同類合作。合作可能帶來的收益（比如好名聲）通常是長遠利益；但意識到有長遠利益可圖并意欲圖之，并不能保證激發追求目標的有效行動，就好比很想長壽的人未必做得到不沾油膩。因而我的假說是，自然選擇為這類行為挑選了一個特殊的動機機制：道德良心。如果你認為某個結果只是你意欲（desire）的東西，你總能對自己說“也許這個欲望得不到滿足也沒有那么糟”。但如果你認為那個結果是值得意欲的東西（認為它具有要求意欲的性質），那么一時沖動貶低其價值的行為就不那么容易合理化了。如果一個人相信某合作行為在道德上是必要的（不論喜不喜歡，他都必須那么做），那么他再跟自己討價還價的余地就小了。如果一個人相信某一行為是某個無法避開的權威所要求的；如果他相信不那樣做不僅會讓自己失望，而且還應該受到批評和譴責，那么，他更有可能做出這一行為。具有實踐影響力的命令的獨特價值就在于它們能阻止進一步計算；在審慎計算可以被各種干擾因素與合理化傾向輕易劫持的情況下，這種阻止是很有價值的。所以我的提議是：從道德角度考慮自己的行為，以一種“我只是不喜歡X”還無法做到的方式，把某些實踐可能從慎思的空間中排除出去，因而自我導向的道德判斷可以充當一種個人承諾。

在此我部分贊同丹尼爾·丹尼特（Daniel Dennett 1995）。丹尼特認為，道德原則的功能是扮演“談話終結者”：在（無論是個人或群體的）決策過程中，道德原則可以用來阻止決策者或決策機制無休止地處理、考慮和要求進一步的理據。“任何一種方針都可能遭到質疑，所以除非我們在設計決策過程時安排一個不用講道理就可以終結討論的方法，我們設計的決策過程將永遠地空轉下去。”（Dennett 1995：506）如果問題是如何處置罪犯，“得到公正審判是他的道德權利！”大概就可以終結討論。如果問題是要不要從商店順手牽羊，“順手牽羊是錯的；我不能那樣做”也可以終結討論。“在我們的世界里這類困境并不陌生”，丹尼特說，“在這樣的世界里，我們不難理解……某種不問緣由的教條主義，保護行動者不陷入過度理性，還是很有吸引力的”（同上：508）。（第五章將進一步討論丹尼特的觀點。）

想象一個沒有道德思維能力的人。假設與某人保持互惠關系對他有利，而他本人，或是因為有意識地想從關系中獲益，或是因為本來就對合作伙伴抱有同情（但也不是無條件的同情），也確實保持著互惠關系。但試想在一次交易中，他因為某個心血來潮的理由（也許他稍一放松就成了眼前利益的俘虜）違背了約定。（這有什么不應該的呢？畢竟，根據假設，他不愿作弊并不是出于原則上的考慮。）現在他會有什么感覺？也許會驚訝于自己一反常態。也許會因為傷害了某個不想傷害的人而失望、痛苦。但重要的是他不會覺得自己有罪，因為只有認為自己違反規范的人才會有負罪感。他沒有道德概念，因此不可能覺得應該為自己的行為受到懲罰。他會后悔，但不會懺悔。出于同情，他或許會想減輕受害者的痛苦（甚至可能想賠償受害者）；但他不覺得自己必須作出補償，因此一旦其他事情分散了他的注意，使他對受害者的同情漸漸淡漠，就沒有什么能讓他再拾起“必須做點什么”的念頭。到頭來，他無非做了件與自己一貫性格不符的、希望自己沒有做的事，僅此而已。“同情”，詹姆斯.C.威爾森（James Q.Wilson）曾說過，“這個脆弱易逝的情緒，來得容易，忘得也快；即便沒有忘記，為袖手旁觀找個借口也很容易。一條流浪狗，一只受傷的雛鳥，都能讓我們心痛，盡管我們知道森林里到處是受傷迷途的動物”（1993：50）。

作為對比，設想有另一個人，不只有同情心還能判斷欺騙是不道德的。縱使她偶然屈從于誘惑，她也會另有一番感受。她能告訴自己說，我犯錯了，我的行為是不正當或不正義的，我必須作出補償，我不僅冒了受罰的風險而且應當受到懲罰。負罪感的情緒對她是可能的。不僅如此，她還能夠做出其他犯人也應受懲罰的判斷（這對我們之前那個未道德化的行動者也是不可能的）。既然這個確實欺騙了別人的道德化的行動者可以對自己做出更嚴厲的指控，那么在她沒有作弊的時候，她不作弊的動機也很可能比非道德化的行動者的更為堅定，也更可靠。盡管以上我對“自我導向的道德判斷能影響行為”這個假說做出了對我有利的解釋（你可以說是書齋空想），但這個假說本身已被大量經驗證據證實。（Keltner et al.1995；Bandura et al.1996；Bandura 1999；Ferguson et al.1999；Tangney 2001；Beer et al.2003；Keltner 2003；Covert et al.2003；Ketelaar and Au 2003.）

需要注意的是，我的論證并不是把未道德化的、受同情心驅動的人，同毫無同情心但有強烈的理性道德責任感的人進行比較（康德的讀者對那種思想實驗很熟悉）。首先，我承認道德化的人可以同未道德化的人擁有一樣的同情心；我論證的只不過是說，道德判斷在原有動機的基礎上有所增益，使得道德化的人更勝一籌。我也不主張道德思維總是優于務實思維；很多時候務實思維是完全堅決的（知道橫穿高速公路是會死人的，恐怕比“亂穿馬路是不道德的”這個判斷更有驅動力）；我論證的只是，務實思維不管用的時候（尤其當實際收益是有變數的遠期利益時），道德判斷可以頂上。另外，讀者必須記住這里說的道德判斷不是康德幻想的那種冷冰冰的、知性化的東西；道德判斷也給一個人的精神生活貢獻如負罪感這樣的情感。我所謂的道德判斷能促發行為的動機，也包括某些道德情感產生動機的效力。“道德情感”一詞在這里可能會造成一點混亂，因為有時它指人們認為道德上值得稱贊的，或總體上親社會的情感（比如仁慈和愛），我們的討論并不涉及這些情感；有時它又指包含道德判斷的情感（比如負罪感）。我所主張的部分內容是，有了后一種道德情感，前一種道德情感更能夠引發行為動機。

初步地說，我的假說是：諸如“那么做不對；我那么做是應受譴責的”的判斷，在慎思、情感以及欲望的形成中扮演著活躍的角色。它能夠促發和強化某些欲望、阻止某些考慮進入實踐慎思，由此提高某些適應性的親社會行為的實施概率。[[58]](#_58_64)支持某一行為的道德判斷當然不能保證該行為的實施，但只要它增加了實施概率，這就有可能是它的演化功能。當然，最終決定一個人行為的是促進某行為的欲望和阻止該行為的欲望的相對強弱。然而最后這點讓我們對前面整個論證產生了疑問。也許道德判斷確實能強化動機，也許在這方面它通常確實優于單純的實踐慎思；但正如生物學家、哲學家大衛·拉提（David Lahti 2003）指出的，還有一個問題有待解決：為什么自然選擇沒有直接加強我們在某些情景中的合作欲望。畢竟，就某些適應行為而言，演化確實賦予了我們很強的相應欲望。比如保護后代的行為似乎就是被單純但強有力的情緒調控的，主要并不是依靠道德責任感。這些情緒基本上不受意志薄弱的影響：幾乎沒有人會為了短期利益去合理化傷害自己心肝寶貝的行為。不僅如此，如果我們的人科祖先已經有了產生如此強大欲望的神經機制，為什么自然選擇這一本性保守的力量沒有利用現成機制來調節新的適應行為，卻產生了一種“截然不同的”且“生物學上史無前例的機制，來實現一個在自然界中經常被更直接的手段實現的目的”（Lahti 2003：644）？這多少有些令人迷惑。拉提的挑戰必須被回應。

每當演化心理學家設想某一適應行為是由一個專門機制管理的，我們總可以問：“自然選擇為什么要費勁去搞那種機制？它為什么不直接創造一種實施該行為的超強欲望？”但考慮一些與道德無關的例子，我們就會看到這個問題其實不太對勁。可以考慮一下人類演化得來的和性與進食有關的心理獎勵系統。我們可以問，自然選擇為什么要費心給我們配備那么多那么復雜的達到性高潮必需的生理機制——為什么不把我們設計成單純地就是想要性生活呢？這個問題是誤入歧途了。自然選擇確實讓我們想要性生活，它用于保證這個欲望的手段之一就是性高潮。類似地，自然選擇讓我們想吃東西，它實現這一點的手段之一就是讓我們覺得食物美味。也許自然選擇也讓我們想要合作，而我們從道德角度去看待合作的傾向（包括產生負罪感的能力）就是它保證這個欲望能得到滿足的一種手段。自然選擇會用特別的手段來創造和加強一類提高適應度的欲望，無論跟道德有沒有關系，這都沒有什么特別神秘費解的。誠然，自然選擇用于加強道德欲望的機制是個“生物學上史無前例的機制”（一種只在人科演化支出現的東西），但如果人類的社會關系確實與其他動物的社會關系截然不同，一個截然不同的解決方法也許是必要的。另須注意，盡管自然選擇是保守的，不同的適應行為常常需要不同的動機機制，仍有一個明顯的理由：如果通過進食和信守承諾能獲得性高潮，我們也許就懶得做愛了。

我們仍可以合理地發問，道德判斷有什么特征，使它得以勝任我們推測它所承擔的演化任務。我認為答案的一個重要部分跟道德判斷的公共性有關。雖然眼下我們關注的是對己的道德判斷，但我們不能因此誤以為我們是在談論私人的心靈現象。正如有公開表達的對己判斷（“我希望你們知道，我對自己的所作所為深感慚愧”），同樣也可以有私下的對人判斷（心中默念“約翰真是個人渣”）。道德判斷，即使是對己的，本質上也是交流性的；事實上，如果我在第三章論證的是正確的（作為語言現象的道德判斷是作為心靈現象的道德判斷的基礎），那么“完全私人的道德判斷”這個想法本身就很有問題。道德判斷可以在集體協商中得到主張，可以用來提出一個要求，證成一個決定，為懲罰提供依據，批評或褒揚他人的行為或性格，或者用來為判斷者自己的人格作證。正面道德評價引起行為動機的方式，與這種公共空間是密不可分的。即便不行騙是因為我個人的良心告訴我“行騙不對”，我也清楚一旦被要求為自己的行為辯護，這種道德考量是可以在公共協商中使用的理據；我也同意，要是我行騙了，他人的懲罰將是有依據的。相比之下，“我就是不喜歡作弊”這一命題可以用來解釋一個人的行為，但缺少道德因素的規范性證成效力（normative justif icatory force）。[[59]](#_59_64)

為了表達清楚起見，接下來我想在另一個稍微不同的辯論背景下闡明自己的立場。按照一種可以說是常識的觀點，日常生活中人們做道德決定的過程是個推理的過程，情緒或動機只是對結論作出合適的回應：據瑪麗的推理，弗雷德的行為在道德上是錯誤的，這個判斷讓她憤怒，然后這個情緒促使她扇了弗雷德一耳光。另一種觀點——其最令人信服的支持者是心理學家喬納森·海特（Jonathan Haidt 2001）——卻認為，常識把事情的順序搞反了。事實真相是弗雷德的行為讓瑪麗感到（比如說）憤怒，這個情緒導致她扇了他一耳光，而她關于他的行為觸犯道德的推理則是一種事后重構。我論證的觀點乍看上去或許和海特相反，但其實不然（這是好事，因為海特提供了大量令人信服的證據，其中一些本章將會討論）。我并沒有說，道德推理促發道德判斷或道德行為。我的假說是道德判斷（不論它是如何達到的——是通過比如推理，還是情緒，還是直覺）對激發行為起了作用。前面說過，道德判斷完全可以被“嵌入”情緒中，所以當我指出道德判斷促發行動時，這也包括負罪、厭惡、贊同等情緒的動機效果。即便許多道德判斷是事后重構的產物，這也不等于說它們在動機方面無所作為。也許弗雷德的行動先促發了瑪麗的憤怒，憤怒的情緒讓她扇對方一耳光，而后她才推理得出“弗雷德是一只性別歧視的豬”這個新的道德判斷。這個新的道德判斷雖然是事后“合理化”的結果，但未必沒有實用價值。也許它比單純的憤怒更能堅定瑪麗跟弗雷德一刀兩斷的決心。更重要的是，它讓瑪麗在公共領域中有一定的說辭來辯護她扇弗雷德耳光的行為，并給其他人提供了反對弗雷德的考慮因素。海特本人也很重視這種聯系，強調說道德判斷“不是在單個人的頭腦中發生的單個行為。它是一個通常十分漫長并涉及多人的持續過程。理由和論證可以傳播并影響人，即便單個人，也很少自己進行道德推理”（2001：828—829）。不僅瑪麗對弗雷德的判斷如此，人們對自身的判斷也是如此。一個人覺得別人會怎樣看待她的決定，她有信心向哪些人證成自己的行為，她是否認為支撐自己行動的理由也能得到同伴的認同（一言以蔽之，她對自己作為社會存在的體驗），將嚴重影響她行動（或不行動）的決心。道德判斷影響動機的方式不是在行為者的腦袋里給某個行為增添一點動力，而是提供一個（看起來是）不能被正當忽視的考慮因素，從而使道德判斷（甚至是對自己的道德判斷）得以在社會舞臺上扮演一個直接的欲望無法證成的角色。當我們認識到這一點時，拉提的問題就解決了。

鑒于以上論證，我要對一開始提出的簡單假說（即道德判斷的演化功能是為某些有適應性的社會行為提供額外動機）作些補充。在道德上否定（而不是單純地反感）自己的行為（或潛在行為），為我們把相應的道德判斷施加于別人身上提供了基礎。不論我有多討厭X，這一傾向本身與我如何判斷他人對X的追求無關。“我不會追求X，因為我不喜歡X”說得通，但“你不會追求X，因為我不喜歡X”就毫無道理。相比之下，“追求X在道德上是不對的”這個主張不但要求我，也要求你避開X。本章開頭我區分了對己道德判斷能帶來什么好處和對人道德判斷能帶來什么好處兩個問題，但我現在要說這兩個問題并不相互獨立（這也不應在意料之外）。當然，我們現在應該看到道德判斷的適應優勢之一正在于它能聯合這兩個方面。道德判斷為評估自我和他人的行為提供了一個共同框架，從而能充當集體協商決策的“通貨”。道德判斷因此可以作為一種社會膠水，將個體的人聯結在一個共同的證成框架中，并為很多群體協作問題提供解決工具。尤其重要的是，盡管未道德化的、強大的負面情緒（比如憤怒）能促發懲罰反應，只有道德判斷才能批準懲罰，因此后者能比前者更有效地管理大群體中的公共決策，尤其是當人們（在一天狩獵采集的勞累后）疲憊不堪，心不在焉，或者不能達成一致的時候。而負罪感（其中包含著認為自己理應受到懲罰的判斷）也能對人有好處，它能抑制人們平時的防御機制，促使他們接受懲罰，或至少做出道歉，從而盡快恢復與同伴的關系（見Trivers 1985：389）。

盡管上文（為了簡潔起見）說道德判斷有加強合作動機的作用，我無意暗示說自然選擇將我們設計成了無條件的合作者。道德感并不是在任何情況下都視合作為善舉的偏好（有這種偏好的人難逃慘遭剝削的厄運）。同樣，雖然我們體內有種先天機制專職促使我們產生進食的欲望，并用愉悅的體驗獎勵我們的進食行為，這并不意味著我們會無條件地吃，不加區別地吃。也許自然選擇為我們設計了非常靈活的合作策略。一個人可以多大方，或必須多吝嗇，將取決于他的環境里有多少資源。誰是有前途的合作者，誰是無賴，這是我們要學的東西。我們可以道德化一個有條件的策略，比如“信賴他人，但不要做冤大頭”。我們也可以把不合作的行為道德化，認為合作在某些情況下是被禁止的。我的主張其實是：將整個靈活的社會結構道德化有助于提高適應性。如此可以避免人們在該合作的時候不合作，但這不應混同于鼓勵人們在不該合作的時候合作。確實，在一個具體場合中，任何對合作動機的增強在某種意義上都意味著人們可能會犯實踐錯誤，比如人們即便在作弊對自己的適應度更有利時仍然去維持原有的交換關系。（想想托馬斯·摩爾的道德信念讓他落得何種下場[[60]](#_60_64)。）但在這種意義上，任何自然的獎勵系統都有可能把我們帶向偶然但可能是災難性的錯誤：人們可能為了果腹而誤食毒草，或者為了床笫之歡而招來強敵。

## 4.3　作為人際承諾的良心

上一節論證了將行為視為道德的要求通常會對一個人的動機有積極影響。當事人由此能得到多少好處則取決于被道德化的行為是什么：如果被道德化的是審慎務實的行為（以上討論的主要是這種行為），將對她很有好處；如果她把對自己有害的行為視為道德的要求，結果則可能是災難性的。然而，還有另一個論證指出，強化一個人不審慎務實的行動決心在某些情況下也可能是對她有益的。既然（我已論證過）道德化的思維是使人堅定決心的思維，這個論證將進一步支持“擁有道德良心有助于個體利益最大化”的觀點。這個論證來自經濟學家羅伯特·弗蘭克（Robert Frank 1988）。

有些顧客買到次品后，會為了討回區區10元退款不惜大費周折。如果你指出與其費這許多周折不如另買新貨，他們對此的回答常常是“你說的是沒錯，但這是原則問題”。如何理解這種看上去非理性的態度呢？有一種說法認為，這種行為可以建立某人不好欺負的名聲，從而減少將來被人欺騙的機會。但這個解答的適用性有限，因為即便在明知眼下的行為不會為將來打交道的人所知時，我們還是經常堅持“原則”。弗蘭克的回答是：這種行為是由情緒控制的（我們可以稱其為“義憤”），而這類情緒的功能是保證某些行為會被付諸實施，甚至不惜壓倒務實的計算，損害自身利益。站在賣家的角度想，如果他面對的是一個總是務實計算的顧客，而他又知道后者退回商品的成本（所費的周折）大于10元，那么他就可以毫無顧忌地向后者出售10元的次品。但換成第二個顧客，他追究起商品問題來不惜成本，會為了討公道做出不理性的犧牲，商家就會確保他賣的10元商品是合格的。現在請你自問是希望成為哪種顧客：不斷被坑的那種，還是獲得良好服務的那種。具有反諷意味的是，后一種顧客（為了討回公道不惜自損的那種）通常不必自損，只要她能把她的“非理性”傳達給賣主，因此就結果而言反而好過前一種顧客。

弗蘭克認為道德良心是情緒官能，其作用與決意退還10元次品的心情類似：確保好的行為得到實施，即便這樣做是非理性的。想想在一項有風險的合作事業中你希望與哪種人為伴：一個基于務實的考慮不斷調整自己行為的人，還是一個能夠承諾合作并會堅持合作，即便當合作從務實的角度看并不理性的時候？卡通人物霍默·辛普森的回答最為精辟：“我們不需要思想家，我們需要實干家——做事不計后果的人！”如果生存下來的前提是你必須被選為合作事業的伙伴（包括被選為配偶），那么成為第二種人是理性的選擇。換言之，如果在你所處的環境中，合作非常重要，那么選取一種能促使你做出有些原本非理性的事情的官能，常常是理性的。那種官能就是良心：它包括了那些，即便是在外部制裁缺失的時候，也能激發符合公認標準的行為的判斷和情緒（其中最值得一提的是負罪感）。（當然良心并不是人們早先自己選擇去擁有的；生物自然選擇已經做了選擇。我的意思是，要是可以選擇，審慎務實的人會去選擇擁有良心。）

弗蘭克的理論促使我們要區分兩種價值：能夠向他人傳達自己的堅定意愿而得到的價值，以及通過執行該意愿得到的價值。前者是價值的主要來源。在10元次品的例子里，價值來自于能夠向賣家傳達自己不惜周折追究到底的意愿。這導致賣家在坑你之前不得不三思，于是你不費周折就能換貨。但如果賣家認為你只是虛張聲勢呢？那么（顯然）你追究這件事的承諾將導致你真的去追究這件事；或許賣家最終同意換貨，但你經歷的麻煩不是10元可以彌補的。但這種情況下，你的承諾不就反過來傷害了你，使得你采取不理性的行動了嗎？未必。弗蘭克認識到，價值還有一個次要的來源：通過執行你說的行動，你保持和強化了，而非削弱了你的承諾，這將有助于你將來的交易。換句話說，某種承諾機制的存在可以讓某個原本不理性的行為變得理性。大費周折去替換一件10元的商品對于一個沒做承諾的人來說也許是不理性的，但對一個已在情緒上承諾了要不惜周折的人來說則未必。這種次要價值不是每次都有的，因為承諾的事也許不僅不便執行，還意味著自毀。（比如揮舞著手榴彈，威脅敵人如果不滿足你的要求就和他們同歸于盡。）在那種情況下，做出承諾確實意味著要進行非理性行為，能夠從中得到的只有主要價值。但在許多其他場合，承諾的存在引入了強化承諾的潛在收益，從而改變了成本—收益結構。為了涵蓋兩種情況，弗蘭克的理論最好不要被描述為“表達一種堅定的執行非理性行為的意愿也許有助于個人利益最大化”，而應該是“表達一種堅定的執行原本非理性行為的意愿也許有助于個人利益最大化”。它給我們的建議不是“魯莽行事”，而是“限制務實計算在你的決策過程中起到的作用”。

弗蘭克對他自己理論的表述有個不足之處，我們現在可以加以改進。弗蘭克認為自己是在解釋良心如何能給其所有者帶來好處，但他只把良心當成一組支持或反對某些行為的、可以傳達的、對動機有影響的感受。［“比如，試想一個能產生強烈負罪感的人。這個人即便在作弊能換來物質利益的情況下也不會作弊。并不是因為她害怕被抓，而是因為她就是不想作弊。”（Frank 1988：53）］但我們已經看到，單純的反感，若不輔以相關的道德認知（例如“如果我這么做了我就是該受懲罰的”），還不足以構成罪惡感，因此也不足以構成良心。盡管弗蘭克成功解釋了為什么自然選擇可能偏愛一個能承諾去做原本不審慎務實的行為，并且還能表達這一承諾的生物，他沒有解釋為什么與單純抑制某行為的強烈情緒相比，把該行為視為“道德禁區”是更有效的承諾工具。

但是我們已經有了給出解釋所需要的資源。前一節的論證提出，通過保障相應行為動機的產生，道德判斷可以充當比單純的抑制更有效的個人承諾。由此可以推論，如果某人能向他人傳達他的個人承諾，那么他的道德判斷就可以充當有效的人際承諾。順著這條思路，我們將稍稍偏離弗蘭克的模型。弗蘭克對良心的理解過于簡單（他只把它看作一組反感情緒），于是他的任務是解釋那些反感如何能令人信服地傳達給他人。他的回答是人類生活中充滿了很難偽裝的、交流性的情感表現（語調、面部肌肉、瞳孔擴張、流汗、臉紅、發抖、笑、哭、眼動、哈欠、姿勢、呼吸頻率等等）。[[61]](#_61_62)我絕不是要無視這些道德中核心的情緒要素，但我認為強調道德的認知要素也很重要。因此我的任務是解釋道德判斷（甚至于那些沒有情感參與的道德判斷）如何能令人信服地傳達給他人。

好在這個任務并不艱難。為了明晰起見，假定我們暫不考慮道德的情感層面。情感絕不是人際承諾的唯一方式。人們也可以通過改變外部限制來做公開承諾——比如（字面意義上的）破釜沉舟，或參與艱苦的成年禮，或簽署有約束力的合同。當丘吉爾向世界聲明“與其讓倫敦卑躬屈膝，我寧愿它化為焦土”，他就是在做這樣一個承諾。承諾的重要之處在于，它排除了未來的某些可能，通過改變他人的選擇（這里“他人”也包括承諾者本人未來的狀態[[62]](#_62_62)）來達到值得欲求的目的。它可以是有意的，也可以是非自愿的（甚至是本性里固有的），可以是絕對的，也可以是或然的。

即便不依靠情感，道德判斷也可以有效地充當這種人際承諾。遵守道德總的來說是件高成本的事，而且其成本顯而易見；遵守道德因此顯然具有充當公共承諾的潛力。（由于這些成本的作用，偽裝道德承諾付出的代價到了一定地步將超過果真愿意履行承諾的代價。）聲明“如果我去追求X，那將是錯誤的行為”就像一個合同，意味著聲明者同意，如果他真追求了X，懲罰他將是適當的。表明道德承諾的方式遠不止口頭聲明這一種。按承諾行事（即永遠不追求X），通常意味著付出代價；如果你確實屈服于誘惑，坦白并接受處罰對你是有害的；大費周折去懲罰其他破壞規范的人可能是有風險的；向子女灌輸這個價值也是個負擔。通過所有這些方式，一個人可以付出一定的成本，向他人發送信號：我承諾用這個道德判斷來指導自己的行為，不去追求X。別人要是接受了此信號，就可能因此調整針對做出道德承諾的人的行為，接受他為合作伙伴：認定他是個值得考慮的配偶人選，是信得過的貿易伙伴，或者干脆就是社會中有價值的一分子。

以上意見應被視為對弗蘭克的理論的補充，而不是對其理論核心的挑戰。毫無疑問，情感在我們的道德生活中確實處于核心地位，情感的展示也很可能是發出個人承諾信號的重要方式。但道德遠不只是情感而已，良心不只是反感而已，人際承諾也不只是意志無法控制的情感流露而已。我認為強調這點尤其重要，因為與恐懼、憤怒、厭惡、快樂相比，自然選擇在設計（無疑是道德良心之核心的）罪惡感時似乎并沒有對怎么以生理表現來傳達這種情感給予太多考慮：研究沒有發現與罪惡感相關聯的獨特表情（Keltner and Buswell 1996）。

4.2節論證了道德判斷可以充當個人承諾，這意味著它們可以完全不必用信號傳達出來。作為補充，這一節論證了道德判斷也可以充當人際承諾，這意味著某種誠實的或有成本的信號是必要的，但我有意淡化了情感信號的重要性。兩種承諾都有點悖論的色彩：要得到承諾的益處，承諾者必須不能以得到該益處為直接目的。就個人承諾而言，籠統而不加質疑的道德思維很多時候比一清二楚的務實計算更有利于個人的現實利益。就人際承諾而言，如果某人承諾去做表面上是（實際上可能也是）徹頭徹尾的不務實的行為，這往往反而對他的現實利益最為有利。（要點在于，發出信號申明承諾的行為可能仍是審慎的，尤其因為如果一切順利，那么發信的結果將是你永遠不必真的做出你威脅要做的不審慎行為。）盡管為了清晰，我把個人和人際承諾分開討論，但4.2節對拉提的回應意味著這兩類承諾在心理上是不可分割的；要理解私下的道德判斷如何讓判斷者下定決心，不能不考慮公開聲明該判斷會如何影響他人的決心。用道德的眼光看待自己，也就是用社會的眼光看待自己，在一個集體的證成框架內評估行為的價值。

至此我們有了一個關于良心的適應價值的合理的假說。這個假說是復雜的，但現象本身就是復雜的，所以這應該是它的優點。

## 4.4　情感的投射

我們之前的問題是“為什么”：為什么自然選擇會產生道德思維這個性狀？但即便一個完整的答案（我確信以上提出的理據還不足夠）仍不能回答下面這個問題：道德思考是怎樣通過自然選擇產生的？最理想的情況下，這個問題應該由神經學和遺傳學來回答。我不能給出這樣的答案，但既然別人也不能，我也問心無愧；連我們最好的科學也還不足以應付這個挑戰。盡管如此，最近的經驗科學研究確實指出了一個明確的方向；這些研究顯示，情感在道德判斷中扮演著中心角色；由此可以推論，如果自然選擇確實直接影響了人類的道德感，它的主要手段是改變大腦的情感架構。（也有證據表明道德思維和長期規劃能力間有聯系，因此大概也為4.2節提出的適應假說提供了一些支持。）以這些經驗材料為起點，本節的主要任務是提出一個哲學上的比喻；這個比喻也許能幫助我們理解史前人類的心靈發生了什么變化，從而獲得了道德判斷的能力。我不打算對“神經層面上發生了什么使得這一切成為可能”這個問題的解答有所貢獻（更不要說基因層面的問題）。

哲學家就情感和理性在道德思想中各自的作用已經辯論了幾千年，最近的科學進展終于取得了實際成果。其中菲尼亞斯·蓋奇的案例堪稱一個里程碑。在1848年的一次可怕事故中，一根直徑為1.25英寸（3.2厘米）的長鐵棒被炸飛，貫穿了蓋奇的頭顱，從左眼下方穿進，從頭頂穿出，這對他的神經系統造成了嚴重損傷（對該案例最好的綜述見Damasio 1994）。讓所有人驚奇的是，蓋奇最終活了下來，除了一眼失明外其他的傷似乎也痊愈了。但人們很快發現他的人格發生了巨大變化：那個曾近有節制、一絲不茍、頗有人緣的人，事后變得粗野魯莽，不負責任，無視道德。在蓋奇之后，與他類似的傷者接受了更精確的測試（Saver and Damasio 1991；Damasio 1994；Anderson et al.1999）。這些受試者患有腹內側前額皮層（ventromedial prefrontal cortex）損傷引起的所謂“獲得性反社會人格”（“acquired sociopathy”），他們的行為（尤其在處理社會事務方面）魯莽不負責任到了不可救藥的地步。他們有自殘行為，無理由地說謊，無法與人建立親密關系，保不住工作，不能為未來制定合理規劃，而且對自己的不端行為和令人發指的無能都沒有表現出絲毫的遺憾或悔意。實驗室條件下一系列模擬現實生活場景的測試顯示，這些人不能為了更大的長遠利益放棄眼前的滿足。情感反應的嚴重衰退也值得一提。（蓋奇本人有沒有衰退癥狀并不清楚。[[63]](#_63_62)）面對會讓一般被試者產生膚電反應的圖片（展示裸體或傷亡的圖片），獲得性反社會人格患者無動于衷。

這些病變研究顯示，道德慎思（以及體現道德慎思的道德行為）與情感能力有很強的聯系。對精神病態者的研究也證實了這個模式；粗略地說，精神病態者的道德障礙與情感缺陷有關聯，但是同理性缺陷沒有關聯（Blair 1995；Blair et al.1997）。這里需要注意，精神病態者的“道德障礙”不局限于反社會的舉止，還包括不能以正常方式（甚至完全不能）做道德判斷。一個正常的三歲兒童能輕易區分違背道德和違背習俗的行為（見下文4.5節），反社會人格者卻不能。更令人印象深刻的證據來自fMRI研究；這些研究顯示，當接受測試的人們被要求在棘手的情景中做道德判斷時，他們大腦的情感區域有顯著活動，尤其是當問題比較“切身”時（比如涉及直接傷害他人的想法）（Greene et al.2001；Greene and Haidt 2002；see also Moll 2002，2003）。對玩“最后通牒游戲”（“ultimatum game”）的被試者的神經成像研究也顯示，當被試決定懲罰背叛的對手時，他們腦部的情感中心變得更為活躍（Sanfey et al.2003；see also Pillutla and Murnighan 1996）。[[64]](#_64_62)

這些證據對“自然選擇如何改變人類大腦使道德判斷成為可能”這一問題給出了一個非常粗略的解答：自然選擇改變了情感中心。這并不是說每個道德判斷都是情感過程的產物，或不存在冷靜的道德推理，或道德判斷不能被理性的方法證成（所以這個解答并沒有削弱道德理性主義），也不是說公共道德判斷的功能是表達這些情感（所以它也不支持非認知主義/情感主義）。[[65]](#_65_62)但我認為這些證據確實在某種程度上支持了一個由來已久的哲學觀點：道德投射主義（moral projectivism）。［這一點上我同意Joshua Greene（2002）。］

對投射主義最雄辯的表述莫過于休謨；休謨指出，心靈具有“把自己擴展到對象上的強烈傾向”［（1740）1978：167］，并主張“趣味”（taste）（與理性相對）“具有創造的能力，它借用內在的情感來給所有自然的物體涂抹或裝點顏色，在某種程度上產生了新造物。”［（1751）1983：88］舉個例子，你見到一只受傷的動物，心生憐憫（pity）。然后，如果你把憐憫這一情感“投射”到這個場景上，那么也許在你看來那只動物就具有一種“要求憐憫”的性質，換言之是可憐的（pitiful）。可憐這個性質，用休謨的話說，是你的心靈“產生”的“新造物”；它看上去好像是這個情景本身的性質，以致你的憐憫似乎只是對這個性質的反應（即沒有參與它的創造），而一個人要是對此無動于衷，毫不憐憫，他就是忽視了一些東西，因而要受到批評。情感的投射并不意味著道德經驗是類知覺的，我們不能像看到紅色一樣看到惡；它意味著我們感知到的世界，似乎都是些不論我們的興趣和態度如何，都束縛著我們的因素，比如殺害嬰兒在道德上不對這件事，在我們看來就是一個基本事實。然而，盡管經驗的表象或許如此（即事件本身的可憐性在先，憐憫之情是其產物），但根據投射主義，可憐性其實是情感的孩子（Blackburn 1981：164—165）。當然，投射主義不是說真的有東西（在字面意義上）被投射了；我們應當將它理解為一個比喻；它想表達的其實是：某些性質表面上似乎是世界自身的性質，但這種表象卻是知覺者頭腦的產物。

人們常假設道德的投射主義和非認知主義互為表里（這兩個詞有時簡直被當成同義詞），但這是錯誤的。假設那些把要求憐憫這個性質投射到世界上的人，要給這個表面性質取個名字，比如稱其為“可憐的”，而后他們到處指著這樣那樣的東西說“這個很可憐”，“那個不可憐”，“那個有點可憐”，等等。至于某一情景為什么在他們看來具有可憐性，即便正確答案是該情景在他們心中激起了憐憫之情，但由此不能得出結論說，當他們說“這樣子真可憐”時，他們的話語的功能是表達憐憫的情感。我在前文明確地論證了，這種表達關系是約定俗成的關系，而不是因果關系，這已足以證明這種推論的無效（2.3節）。事實上，投射主義和非認知主義連和睦共存都很困難。這里需要注意的關鍵之處在于，投射主義蘊涵著一個關于世界在投射者們看起來是什么樣子的理論：在投射者看來，世界包含某些性質。既然我們能假設他們用以談論事物的語言反映了他們的經驗，那么當他們說“那個受傷的動物很可憐”時我們可以假設他們是在斷言那個情景里的事物具有可憐性。但如果他們確實如此斷言（也就是說，他們在表達對此事的信念），那么他們不可能只是在表達情感。

在休謨看來，投射主義的適用范圍遠不止道德。投射主義也是他最青睞的關于聲音、顏色、冷熱以及因果性的（著名）理論。他知道這是違反直覺的——“在常人看來，哲學中迄今為止提出的最大悖論莫過于雪不白不冷，火不紅不熱”［（1762）1986：416］——然而，鑒于前一段所說的內容，投射主義必然是違反直覺的。舉例而言，關于冷熱的投射主義，揭示了這些特質在我們看來是什么樣的：它們好像是客觀世界的一部分，是火和雪的真實性質。因此任何揭露冷熱的投射本質的哲學觀點從日常的觀點看都必然是離譜的；因此根據假設，投射主義就是違反直覺的。

“惡習和美德”，休謨說，“就猶如聲音、顏色和冷熱；按照現代哲學的觀點，它們都不是物體的性質，而是心靈的知覺”［（1740）1978：469］。休謨援引的正統觀點，是受伽利略、牛頓和洛克（這些人都被理解為顏色投射主義的支持者）啟發的18世紀投射主義理論。投射現象是否像休謨認為的那么普遍存在，這一點可以商榷。（對顏色投射主義的現代辯護，見Boghossian and Velleman 1989）。但我確實認為演化科學至少增加了投射主義的可信度，因為它給我們解答了一個把實踐成就（而不是準確性）視為最高價值的過程是如何設計了我們。在生活的很多方面，顯然準確是獲得成功的最佳途徑：對于樹上有沒有果實、有多少、在哪個方向，最有用的信念就是正確的信念。但在其他領域（特別是關于顏色、聲音等感覺形式）犧牲一點真實也許能換來更大的利益。要更好地從視覺上區分成熟和未熟的果實，也許感覺經驗可以夸大兩者的差異。事實似乎正是這樣。紅和綠在我們感覺意識中的差異比現實中對應的光頻率的差異顯著許多。這幾乎反映了我們的祖先曾經常吃果實，那個時候識別果實的成熟度可是頭等大事（Johnston 1999：14—15）。

自然選擇為什么要用這些投射傾向來拖累我們呢？最吸引人的答案是：這完全不是“拖累”。正相反，這似乎是設計有知覺的生物的最高效方式。首先，考慮一種沒有被投射到外部世界的知覺性質：疼痛。你用叉子戳自己的腿，隨之而來的疼痛是作為某種屬于心靈的東西出現的；你一秒都不會認為疼痛是叉子的性質，只在突然接觸的瞬間傳到腿里。疼痛沒有被投射這一事實想必要用疼痛的意義來解釋：疼痛的全部意義就在于告訴生物，他們的身體出了問題。最重要的是我在痛，而不是叉子具有制造疼痛的性質。但試想一下，如果我們所有的感官知覺都是被如此呈現的，結果會多么奇怪。在這種情況下，你對一團通紅灼熱噼啪作響的火的經驗，將不再是對于具有這些性質的、真正的火的經驗，而是對于你自己的（由外界性質導致的）熱的感覺、（由外界性質導致的）紅色的感覺和（由外界性質導致的）嗶剝聲的感覺的經驗。我認為這種可能性毫無自相矛盾之處（不難想象也許某個生物對世界的知覺經驗就是那樣的），但顯然這不是事物向我們呈現的方式。原因是明擺著的：這種體驗世界的方式在繁殖適應度上沒有任何優勢。更簡單且完全適當的方案是將我們的經驗呈現為對世界的經驗：對火的通紅、灼熱、噼啪作響的經驗。在同等條件下，最簡單的設定是讓生物的知覺經驗成為對世界的某些方面的親知（direct acquaintance），仿佛感官不過是開向現實的窗戶，仿佛事物恒常的表象就是它們的本來面目。就尋找成熟的果實、躲避豹子、找到并打動未來的性伴侶、撫養孩子、制造工具等目的而言，感覺經驗如果總是標明自身的心靈現象本質，就是毫無意義的浪費，分散精力的障礙。從這個觀點看，投射的傾向遠不是浪費，可以想見，它正是嚴格高效的自然選擇的產物。

當然，以上遠不是完整的論證；相關哲學辯論之復雜令人生畏。我無意就休謨的雪不白、火不熱等觀點下任何定論，我只想指出我們對自然選擇的理解給投射主義提供了一些初步（prima facie）的支持。但就道德的投射主義而言，我認為可以更進一步。我在此要澄清，我并不是說現有的證據足以證明道德投射主義（即我并不是在論證現有的證據足以排除所有其他理論），我只是想論證說，現有的證據與投射主義的解釋十分契合，因此把投射主義采納為工作假說，也許可以幫助我們理解我們祖先的大腦發生了什么變化，使其獲得了道德判斷的能力。

什么樣的經驗證據能支持投射主義？首先它必須能支持一個道德現象學的論斷：道德屬性看起來是“存在于外部世界之中”的。其次它必須能支持一個發生學的論斷：這樣的道德表象實際上主要是由情感活動導致的。由此衍生的一個推論是，這種表象在某種程度上具有欺騙性；實際上是情感活動促發了我們的道德判斷，但我們卻覺得情感是對世界本身具有的屬性的反應。此外，或許還要加上一個反面要求：沒有證據表明，道德感的功能是發現真正的道德事實。要是經驗證據滿足了以上這些要求，是否就足以證明投射主義，我（如前所述）表示懷疑；但這絕對能使投射主義成為一個有競爭力的理論。有些證據我們已經看到了；另一些我先簡單地提及一下，以后再詳細討論。

首先，投射主義所需的道德現象學論斷似乎不乏支持。研究顯示，“常識道德”確實包含一些關于道德客觀性的主張。有研究者（Nichols and Folds-Bennett 2003；Nichols 2004）調查了幼兒對“太甜”“好吃”“無聊”等性質的反應并與他們對道德和審美性質的態度做了比較。下面舉兩個那些4到6歲的被試者被問到的問題：

想想很久以前，還沒有人的時候。那時候還是有葡萄的，跟現在的葡萄一樣。在那個時候，在人類出現之前，葡萄好吃嗎？

我覺得猴子之間互相幫助是好事。有的人不喜歡猴子在受傷的時候互相幫助。他們覺得猴子那樣做不好。如果一只猴子受傷了，另一只猴子來幫它，你覺得這只是對于某些人來說是好事，還是它本來就是好事？

被試兒童將所有性質的例子都視為獨立于人類的存在（即他們認為還沒有人的時候葡萄也是好吃的，玫瑰也是美的，等等），但他們會明確區分基于和不基于偏好的性質：好吃的東西只是對某些人來說好吃，過甜的東西只是對某些人來說過甜，但好事“本來”就是好的。綜覽此類證據，肖恩·尼克爾斯（Shaun Nichols 2004：176）得出結論說：“鑒于來自幼兒的證據，我們應當把‘道德客觀主義是常識元倫理學的默認設定’理解為一種有效的假說。”[[66]](#_66_62)這些結果與發展心理學中另一大類研究文獻相輔相成。后者顯示，人們不論文化背景，都傾向于把獨立于權威的違規行為和對某人旨意的違反區別對待。（4.5節將討論這類文獻。）拉里·努奇（Larry Nucci 1986，2001）甚至發現在門諾會和阿米什[[67]](#_67_60)的兒童和青少年看來，連上帝的權威也不能決定道德上的對錯。被問及如果上帝說禮拜天可以工作，那么禮拜天是不是可以工作，所有人都說“是”；但被問及如果上帝說偷東西是可以的，偷東西是不是就是可以的，超過80％的人說“不是”。這些結果，和其他令人信服的經驗證據一起，表明在人們的經驗里，道德命令和價值是“客觀”的，即在人們看來，道德命令和價值既獨立于我們，也獨立于任何權威。

其次，投射主義所需的道德發生學論斷似乎也不乏支持。我已經提過來自反社會人格、病變研究以及神經影像的證據，這些證據都表明道德和情感官能之間有聯系。不僅如此，經驗研究還表明了這種聯系的走向：是從情感到道德判斷。喬納森·海特搜集了大量證據來支持這一假說（尤其參見Haidt 2001）。在其中一項研究中（Wheatley and Haidt 2005），被試者被深度催眠，然后每當他們讀到某個任意的單詞（“often”或“take”），都被暗示產生厭惡感。隨后，實驗人員要求被試者閱讀各種虛構的小故事（有的故事涉及違反道德的行為，有的不涉及任何違規行為，有的使用目標單詞，有的不用。）結果，被促發厭惡感的被試者的道德譴責也加強了。甚至在毫無違規行為的故事中，感到厭惡的被試者也常常傾向于追加負面的道德評價。這結果相當驚人，因為故事本身完全不能支持這種判斷：

丹是校學生會的成員。這個學期他負責為關于學業問題的討論制定議程。為了激發討論，他試著挑選（tries to take）或通常選擇（often picks）教授和學生都感興趣的話題。[[68]](#_68_60)

感到厭惡的被試報告說“他好像圖謀不軌”，“他是個迎合別人的勢利小人”，“看上去就很奇怪，很惡心”，或者“我也不知道（原因），但反正就是不對。”這些證據表明，很多時候是情感在驅動我們的道德判斷，盡管我們主觀上意識不到。海特的研究多次表明，通常人們不知道促發他們道德判斷的真正原因；人們上下求索理由和根據，費盡心思以避免承認“我之所以認為它不對，是因為它讓我覺得憤怒，僅此而已。”

這些（一方面）跟道德判斷的真正起因以及（另一方面）我們的主觀體驗有關的證據，實際上已經構成了對道德投射主義的表述。沒有證據表明人類的道德感像知覺器官一樣，能夠發現世界本身的道德性質。到目前為止，本書對于道德判斷為何會有適應性的討論，既完全沒有要求，也絲毫沒有道德判斷的功能是描述客觀道德事實的意思；更沒有任何暗示說自然選擇為道德判斷設計的功能跟發現世界的性質有什么關系，相比之下，其設計功能更像是鼓勵成功的社會行為。但該假說同時也指出，道德判斷若要實現這個功能，它們必須看上去像是確實在描述客觀的道德事實（至少在“讓實踐考慮具有無法逃避的權威效力”這個意義上）。（如果你只把“談話終止器”當成談話終止器，它就不再能有效地起作用了。）所以，4.2和4.3節提出的關于道德判斷的適應性功能的演化假說本身就暗示了道德投射主義。在該假說描述的生活領域中，自然選擇并不關心人們其實發現了什么，但一個類似于發現的主觀體驗卻是至關重要的。

道德投射主義還有望幫助我們理解一個原本費解的現象：無法解決道德爭議。如果做道德判斷就是發現事實，我們豈不應該能達成某種共識？就算不能在所有問題上達成共識，至少也應該能就解決爭議的程序達成共識。但道德爭議可不是一般的頑固難解。我在此大膽主張：任何人曾經做過的任何道德判斷，都會被另一個智力完全正常也熟悉相關情況的人認為是錯誤的、有害的、偏頗的、狹隘的。驚人的是我們對解決這類爭議的方法也沒有共識。美國總統可以自信地談論“邪惡的存在”，然而沒有人能告訴我們邪惡是什么性質，用什么方法發現它的存在，這一論斷的真偽又如何驗證。于是，持續而無望解決的爭端隨處可見：“擱置爭議”通常就是我們可以期望的最佳結果。但即便某種程度的寬容得以實現，爭議雙方通常仍會認為對方深陷于錯誤的意識形態和（或）自私中不能自拔（同時他們自信自己看事情既清楚又正確），而我們這些幸運的旁觀者通常則認為爭議雙方其實半斤八兩，都深陷在這些問題里。如何理解這個現象？也許道德判斷的任務的確是探尋道德事實，情感只是搗亂而已；但在我看來，對于激烈而徒勞的道德糾紛，以下的解釋可信得多：情感才是主宰，道德判斷是情感被投射到個人的社會經驗的結果（Haidt 2001：823）。換言之，雖然我們可以想象“只要（假設不可能為可能）能杜絕情感，我們的道德判斷將變得更清晰更正確（也將達成更多道德共識）”，但事實似乎更可能是：倘若沒有情感，我們再也不會做出任何道德判斷。

雖然我一直在支持道德投射主義，但我的意思不是說人們做的每一個道德判斷都是某一段情感的產物。我的主張主要是圍繞譜系展開的：在人類進化的歷程中，投射官能的出現是解釋人類道德判斷能力的重要一環。這個譜系性的主張并不排除我們在某些場合（也許經常）通過其他方法做出道德判斷的可能（見Haidt 2003b）。顏色投射主義理論也是同理。假如我要你說我家門廳的地毯是什么顏色。你說不出來？好吧，給你一個重要線索：它的顏色跟成熟的西紅柿、血、覆盆子差不多。這下你應該不難猜到我家的地毯是紅色的。你做了一個特定的顏色判斷但沒有使用“休謨式投射”。盡管如此，這個顏色判斷很明顯在某種意義上仍然依附于“典型的”顏色判斷，即當你看著合適的對象（比如西紅柿），并且有正常的對于紅色的主觀視覺經驗時所作出的判斷。真正包含投射的（休謨主義者會說）是這種典型顏色判斷。我認為，道德投射主義者也應該承認類似的東西。確實，有些（真誠的）道德判斷不伴有情感活動。但在某種意義上（在此我不打算說明是什么意義）這些判斷依附于有情緒經歷伴隨的典范判斷。如果沒有人做典型顏色判斷，就不會有人做任何顏色判斷；同理，如果沒有人做包含情緒投射的典型道德判斷，就不會有人做任何道德判斷。[[69]](#_69_58)

開篇的問題是自然選擇如何帶來道德判斷能力。我們的大腦發生了什么變化，讓我們能用義務、公平、應得、財產、作弊等概念來思考？我認為這種變化可以在投射主義的框架中得到合理的解釋。如我們前面所見，單純的好感和反感不足以構成道德思維；討厭一個結果與不允許那個結果的出現大不相同。我們需要的是一個從欲求（鄙視、贊揚、拒絕、要求）一樣東西到認為那樣東西值得欲求（值得鄙視、值得贊揚、不可接受、被要求）的轉變。解釋這一轉變恰恰是投射主義的長項。而一旦有了諸如可欲，可鄙，值得贊揚，不可接受，被要求之類的概念（順便提一下，只有使用語言的生物才能有這些概念），一類看上去不容易用投射主義解釋的規范概念也可以解釋了：可欲但不被要求，不適當但未被禁止，英勇、吝嗇、冷酷、正義、墮落、好客、無情、粗魯等等。如果投射傾向，因其拓展一個人的概念空間的潛力，可以被稱為“知性能力”，那么我們也終于可以理解達爾文令人困惑的（前文3.7節引過的）論斷：親社會情感知性化的過程就是道德化的過程。（不過我無意宣稱這就是達爾文本人當時的想法。）

按理說，我們還可以追問“但自然選擇是如何帶來這個投射傾向的？”（神經層面上如何？基因層面上如何？）這是個好問題，但它超出了我們現有的能力。不過，手頭缺乏一個完整的令人滿意的理論并不是件稀奇的事：相關領域里到處都是懸而未決的問題。自然選擇如何（在神經和基因層面上）造就了人類的語言官能，我們對蛇（而不是對槍）的天生的恐懼，我們區分男女面孔的能力，乃至像我們對性的興趣這樣基本的東西，我們對這其中任何一個問題都沒有完整的答案，但這個事實應該激勵進一步的研究，而不是助長對這些現象是否是自然選擇直接產物的懷疑。

## 4.5　無非是個“原來如此”的故事？

上面的故事聽起來還不錯。但它和被史蒂芬·杰伊·古爾德（Stephen Jay Gould 1978）炮轟的“原來如此”的故事（這一稱號來自吉卜林講的豹子如何得到斑點、大象如何得到鼻子的故事集）有什么不同嗎？我必須承認我不喜歡古爾德的用詞，而它在坊間的流行，無非使我更加反感。首先，吉卜林的故事是純粹的幻想（與嚴肅合理的解釋性假說有天壤之別）。如果這里的抱怨僅僅是它還有待驗證，那就直說好了，何必冷嘲熱諷呢？其次，要記住古爾德的指責主要針對的是“適應主義的宏大計劃”（the“adaptationist program”）：這種觀點認為，生物的所有性狀都是完美的適應性狀，因而唯一的問題是性狀如何演化，而不是它們是不是演化而來的。有沒有人真的持有古爾德攻擊某事這樣的極端觀點，就留待別人去判斷。（關于該辯論的歷史細節，見Segerstråle 2000。）為了明確起見，我在這里要指出，我并沒有假設所有生物都具有完美的適應性，也沒有假設所有性狀都是適應性狀。我認為“人類道德不是被選擇的性狀”是個值得嚴肅考慮的假說。或許道德就類似于古爾德所說的拱肩，只是自然選擇的偶然的副產品，本身并無演化功能（Gould and Lewontin 1979）。或許它是丹尼特（1995：77—78，485—487）說的“好把戲”：道德的好處顯而易見，因此我們可以認為任何一群智力中等的人類都會創造它。雖然我認真對待這些假說，但我認為他們不太可能是真的（我這樣認為并不是因為我喜歡聽上去合理但沒有經驗根據的、不能驗證的童話，而是證據使然。）我們必須認識到在這類事情上永遠不會有一個以明確為真的前提出發的演繹論證；我們永遠不會得到也不該指望百分之百的確定性。我擔心的是，在某些圈子里，“故事‘不過如此’”這個反駁變得太習以為常，以至于人們會用它來攻擊任何根基尚淺的譜系解釋性假說。若果真如此，我們大約會繼續聽到古爾德的經典嘲諷被用來攻擊完全值得尊敬、得到充分驗證的科學理論。

我以為，要免于合理的古爾德式責難，假說必須是可以檢驗的。一個假說也許目前還沒有被檢驗，但那不是對假說嘲諷譴責的理由。關鍵是我們對它的態度要恰如其分，只把它作為還沒有得到檢驗的假說來看待：就同在其他科學領域一樣，這里把聽上去合理但未經驗證的假說信以為真是錯誤的。古爾德說，如果在某個領域，“發明的精妙取代了可檢驗性成為是否接受一個理論的標準”（1978：530），那個領域就出了問題，這種說法是正確的。鑒于此，讓我明確聲明，我的假說不是作為真理被提出的。它是個合理、融貫且可以檢驗的假說——但它的真實性還有待驗證。不過，我們也有充分的理由對它青眼相看。在本章的最后一節，讓我再回顧一些支持它的證據，完成本書的第一個任務。

道德（在此我指的是做出道德判斷的傾向）存在于迄今所知的一切人類社會。古埃及《死者書》和美索不達米亞的《吉爾伽美什史詩》都提到了道德戒律。如果貿易意味著對財產的理解，而財產意味著某種對權利的理解，那么考古記錄中道德的實體遺跡還可以上溯很遠，至少到舊石器時代晚期（Mellars 1995：398—400），也許還更早（McBrearty and Brooks 2001）。沒有任何證據表明道德（或尤其是貿易）是從某一個或某些個文化創新中心起源和傳播現代文明的全新發明。相反，像語言一樣，它普遍而古老。

此外，幾乎每個人都有做道德判斷的傾向。它的發展無須正式指導，無須有意努力，而且個體對它的特點也沒有自覺地了解。比如，年紀很小的兒童就能區別道德規范和務實規范（Tisak and Turiel 1984），盡管沒有人明確教過他們，因為大多數成年人也沒有明確意識到這個區別（甚至許多努力思考這類事情的哲學家也沒有）。這不是說兒童必然會發展出道德判斷；就像習得語言一樣，與某些環境特征的接觸是必要的。但所有證據都表明嬰兒的大腦已經為道德判斷“做好了準備”，就像鳥類的大腦為學習鳴唱做好了準備一樣[[70]](#_70_58)（Seligman 1970，1971）。道德在人類幼兒中的發展開始得非常早，發展順序極為穩定，發展路徑獨立于其他技能的出現，并包含突然成熟的過程。關于最后這點，海特（2001：826—827）如此描述人類學家艾倫·菲斯克（Alan Fiske 1991）的發現：

四歲以下的兒童對公平問題顯得相對不敏感，但在四周歲左右時，他們對公平的關心驟增，并且擴展到對公平的關切從未鼓勵，而且通常也不合時宜的社會領域。這個模式（在類似時間突然出現的擴展）向我們顯示，這些現象來自一種內源（endogenous）能力的成熟，而不是對一系列文化規范的習得。

我說的“道德發展”不只包括親社會行為，更不只包括親社會情感；我指的是真正認知性的（盡管也許仍然嵌入情緒的）道德判斷。大量發展心理學研究展示了這個性狀產生時的方方面面。首先，我們知道部分兒童早在三周歲時，已在直陳條件句和道義條件句的處理上表現出和成人一樣的分離效應。舉個例子，面對一個房子模型和一些玩具老鼠（有些老鼠捏了會叫，有些不會），學齡前兒童不能正確指出，為了檢驗直陳條件句“會叫的老鼠都在房子里”是否為真，他們要捏哪些老鼠（Cummins 1996a）。但當內在邏輯相同的任務用道義語言表述時（孩子被問到為了檢驗“會叫的老鼠都必須待在房子里”這個規則，需要捏哪些老鼠），他們的表現好了一倍。（Harris and Núñez 1996的研究支持了這些結果。）對成年人的類似測試顯示這個效應完全是跨文化的（Cummins 1996b；Sugiyama 1996；Sugiyama et al.2002[[71]](#_71_56)）。

幼兒不僅在與道義相關的事務上表現出更優秀的推理能力，而且很早就能區分不同類型的義務規則。[[72]](#_72_56)最值得一提的是，區分道德違規和習俗違規的能力早在第三年就出現了（Smetana 1981；Smetana and Braeges 1990），而且這個現象具有驚人的跨文化普遍性（Nucci et al.1983；Hollos et al.1986；Song et al.1987；Yau and Smetana 2003）。甚至在貧困和虐待中長大的兒童也能區分道德和習俗（Smetana et al.1984）；事實上，不能做出這種區分的大概就只有精神病患者（Blair 1995），但他們的特征和成長經歷基本沒有關系，而幾乎肯定是神經病變的結果（Wootten et al.1997；Laakso et al.2001；Kiehl et al.2001）。

研究道德/習俗區分出現的學者們，大都借鑒了心理學家埃利奧特·圖列爾提出的以下幾個區分維度：違反道德被認為比違反習俗更嚴重，更具普適性（比如，在別的國家也是錯的），反對的理由不同（比如，反對違反道德的理由與有害的后果有關），并且道德違規被認為是獨立于權威的（Turiel 1983，1998；Turiel et al.1987）。這四個條件是否總是抱團出現或是否構成一個自然類，或許可以懷疑（Shweder et al.1987），但圖列爾等人的發現無疑是一個發展過程中的深刻區別。特別有趣的是最后一個，必須動用反事實思維來理解的維度。就習俗違規（比如男孩穿裙子上學）而言，當被問及“如果老師說這沒問題，那會怎樣呢？”，兒童會說原來的規則不再具有約束力。就道德違規（比如毆打另一個學生）而言，不論老師怎么說，兒童都傾向于堅持它是錯的。（不到三歲的兒童也顯示了這一分離效應（Smetana and Braeges 1990）］。這一觀察與我關于道德判斷的特殊的“實踐影響力”的主張（以及前一節提到的Nichols和Folds-Bennett實驗）契合。

這些來自發展心理學的結果強有力地支持道德判斷傾向的先天論。不可否認，文化學習在決定一個人的道德判斷的最終內容上起著重要作用；但我主張的是，自然選擇設計了一個（或一系列）特化的先天機制，使得這類學習成為可能。上述發現的跨文化性應該給我們留下深刻印象。跨文化共性并不蘊涵先天性，先天性也不蘊涵跨文化共性。然而，一方面，只要沒有環境變量制造差異，先天特征傾向于導致跨文化性狀。另一方面，跨文化共性也需要解釋；因此每當我們發現一個跨文化共性，我們必須問有沒有合理的替代假說解釋。我認為反對道德先天主義的人（比如支持丹尼特的“好把戲”假說的人），在解釋這些跨文化結果時會面臨很多困難。他們的理論是要求千姿百態的人類社會中始終存在一個穩定的外源（exogenous）因素，該因素可以合理地解釋兒童道德發展特有的，極其規律的個體發生順序。然而，這種因素存在的可能性微乎其微。這里我的論證邏輯是最佳解釋推理（inference to the best explanation）；就目前而言，先天性是最佳解釋。

這個論證也受到以下考量的支持：很難想象，一般性的學習機制怎么能夠學會道德判斷。這里的問題不只是環境沒有提供足夠的材料讓兒童來掌握必要的區別；問題是，我們甚至不清楚在特定環境里面（哪怕一個豐富多變的環境中），到底會有什么樣的東西，可以讓一般性的學習機制采納并以此發展出道德違規概念。20世紀中葉的社會學習理論者認為懲罰是把規范內化的途徑，他們也許是對的，但他們并沒有回答，懲罰除了讓受罰的人對懲罰產生畏懼，對會遭受懲罰的過錯產生反感之外，如何還能有別的結果。你可以盡情懲罰一個具有一般智能的東西；光憑這個，它什么時候能學會道德違規這個概念？說“行為和懲罰之間的關聯被內化了”只是給問題取了個名字，而沒有解釋任何東西。其他的發展理論在這一點上也沒有好到哪里去。他們要么訴諸“內化”，將后者當成某種人類心靈恰好具備的神秘活動（仿佛這個傾向本身可以通過觀察和運用一般性的智能從環境中學到），要么就是暗中預設心靈生來就為內化規范“做好了準備”。

以上關于道德判斷如何在個體層面發生的問題，反映了一個任何關于道德的人類史前文化起源的假說都要面臨的挑戰。這個挑戰在盧梭論述人類文明基礎的著名段落中暴露得相當明顯：“第一個把一塊地圈起來說“這是我的”，并發現人們單純到竟然相信他的人，是文明社會的真正奠基者。”［（1758）1997：164］如果盧梭想象的是人們對占有（possession）的概念已經有了堅實的理解，只不過沒有想過擁有的是土地，那么他的話里（至少就我們的目的而言）沒有值得評論的東西。但假定盧梭想象的不是土地所有權的起源，而是所有權這個概念本身的起源。設想他的神話主人公劃出一塊地，然后向同胞宣布：“這是我的。”然而，如果他的同胞對所有權沒有任何概念，他們對“我的”是什么意思也不會有概念。[[73]](#_73_56)他們會回答：“你在說什么？你說你和這塊地間是有什么的特殊關系？”那主人公接下來怎么辦？他如何能向連所有權也不懂的人們解釋占有一個東西意味著什么？他也許可以說：“我對這塊地有特殊的，你們沒有的權利。我可以隨意處置這塊地，對此你們有義務尊重，而如果你們隨意處置它，我也有權反對。”但是根據假設，其他人不但不理解所有權，也不理解道德權利和義務，乃至任何道德概念，所以這類解釋也不太可能奏效。就算他們觀察入微，推理嚴格，只要他們還沒明白道德思維，無論你說什么，做什么，強調什么關聯，實施什么懲罰，或者鼓勵什么推理，都不可能理解“這是我的”是什么意思。

這里我們有必要暫停一下，再考慮考慮上面列舉的最后一項，因為圖列爾等人似乎認為兒童是通過推理學會區分道德和習俗的。但是對比（比如）故意傷害的行為和男孩穿裙子的行為，有什么東西能讓一個觀察新手推斷出對前者（而非后者）的禁止是獨立于權威的，這問題其實很困難。要知道，為了推出依賴關系，人們必須觀察到相關權威改變主意允許男孩穿裙和該行為不再違規之間存在著關聯。而要推出獨立關系，觀察者必須發現（A）盡管相關權威改變了它對某傷害行為的觀點，該行為仍算違規，或者（B）之前被譴責的傷害行為不再被認為是違規，但相關權威的意見并沒有改變。但這兩種現象即便成人都難得見識，更不要說三歲小孩。對于嚴重的道德過錯，比如暴力犯罪，我們所見無一例外是兩個因素都穩定不變：所有相關權威都譴責它，它也繼續構成違規。一個人要怎樣才能由此推出獨立關系，實在令人費解。兒童的確可能發現違規行為的嚴重程度參差不齊（有些讓人更憤怒，受到更重的懲罰），但是在他們體驗到的外部世界中，并沒有任何東西能幫他們將這個可觀察的區別，同依賴于權威和獨立于權威的違規行為之間的、通常不可觀察的區別對應起來。再說，什么經驗能讓兒童推斷出哪些規范是地方性的，而其他規范更有普適性？如果適用范圍區別僅僅是，比方說，學校和家的區別，我們尚能在兒童的經驗中找到區別的根源。但許多習俗在校在家都適用，而且就相當一部分社會規范（比如，進餐須用餐具，不能手抓）而言，兒童對于不適用這些規范的環境其實通常既沒有直接也沒有間接經驗。此外值得一提的是，圖列爾等人似乎認為與同伴的互動為兒童提供了大部分推理的基礎（Turiel 1998：899；see also Nucci 2001：13—16）。但這沒有回答問題，只是把問題往后推了一步：其他兒童是如何學會這個區分的？總而言之，“世界各地的兒童通過基于某種相似互動經驗的推理而學會區別道德和習俗”這個假說，據我所知從來沒有被詳細闡明，而且細想之下根本難以置信。（類似的想法見Dwyer 1999；另見Sripada and Stich,forthcoming。）

解決疑難的關鍵，是認識到道德不是兒童從外部環境中學到或推導出來的東西，也不是我們的祖先在史前迷霧中憑著一般性的才智發明的東西。只具備一般性才智的生物不可能發明或被教導學會道德判斷，不可能學會把反感轉化為不許可，因為明白道德判斷是怎么回事需要某種特定的大腦：一種為此類學習配有專門機制的大腦，而這種機制正是特定演化過程的產物。這個演化假說遠不是一個“不過如此的故事”；它看上去是我們現有的最好的故事。

## 4.6　結論

就像不存在呼吸基因一樣，也不存在道德基因。我們的大腦也沒有哪塊地方像過去骨相學家想的那樣，是專門用來做道德判斷的。道德經常也是相當復雜而模糊的事，而道德判斷無疑涉及多種心理和神經機制。但這些對簡單化思考的警告，并不是在否定本書提出的假說：在我們這個演化支中，自然選擇了道德思維。另外，我之前強調過，雖然道德思維對環境中的微妙細節極為敏感，這個事實也不削弱本書的假說。文化的傳遞機制在決定一個人的道德信念內容方面發揮著巨大甚至獨一無二的作用。但這并不排除可能有一種專門為了使這類文化傳遞成為可能而設計的先天“道德感”。

生物學有沒有對個人或一個文化的道德框架內容做任何限制呢？也許是有的。2.5節提到跨文化研究發現了一些各道德框架的共性。這些研究強有力地顯示，人類的道德感天然地傾向于關注環境中的某些因素而非另一些因素，比如故意傷害的事件，互惠關系的保持（公平、作弊等），社會地位，以及一些關于身體和身體功能的主題。當然，這些道德范疇的普遍性本身并不意味著產生道德的傾向是先天的；但綜合考慮所有證據（包括前面討論過的來自發展心理學的證據），對道德的普遍特征的最佳解釋似乎是：它們是人類心理的基本“設計特征”（“design feature”）的表現。至少存在一個非常合理并且可以檢驗的類似假說，這一點必須承認。

鑒于以上，現在可以就一個我迄今都回避但某些讀者可能最感興趣的問題稍作評論：第一章概述的諸多演化過程中，哪一種是人類道德感的成因？我回避了這個問題，無非是因為我們不知道答案。本章討論的經驗證據與第一章提到的每個過程都可以融貫一致，但是我們不知道要去哪里尋找堅實的證據來解決這個問題。[[74]](#_74_56)盡管如此，還是有個引人注目的事實：所有的人類道德系統都把互惠關系放在主導地位；如果人類的道德感是為了某一件特定的事情而準備的，那絕對是互惠關系。因此我們可以相當合理地猜想互惠交換是道德被設計來解決演化上的一個核心問題。這么說并沒有排除其他過程的可能性。群體選擇（多半是文化層面上的）也完全可能是一個主要因素。但我的感覺是，（廣義上的）互惠是一切的起點。（至于厭惡的道德化——由此產生了關于比如食物和性的禁忌——我懷疑它是自然選擇借用了為其他目的演化而成的動機機制的結果。）來自靈長類學，實驗經濟學、神經科學、發展心理學和人類學的證據都顯示，人類的心靈還帶著某些來自過去的印記，標志著互惠曾經扮演過重要的角色。我們對別人名聲的好奇，對宣揚自己美名的熱衷、對交易公平的敏感、區分偶然和故意傷害的能力（及原諒前者的傾向），對欺詐的敏感和反感（及不惜自己付出實質代價也要懲罰欺詐的意愿），以及高度的占有意識——這些很可能是先天的傾向，這都表明，我們的心靈是為了互惠而生的。

盡管在此我將不再發展這個互惠假說，但有兩點值得一提，以防誤解。第一，有人也許會說當今很多道德實踐跟互惠沒什么關系：我們對子女、殘疾人、后人、動物以及（如果你樂意）對環境的責任，似乎都不依賴于我們對回報的期待。但這個反對不在點子上，因為這些事實完全無礙于“道德最初是為了調節互惠交換而演化出現的”這個主張；我并沒有宣稱今天繼續維持社會關系的只有互惠而已。導致某人撲到手榴彈上以拯救同伴的責任感也許是拜互惠所賜。這里沒有實際的互惠（甚至沒有對此的期待），但促發自我犧牲的心理機制仍可能是互惠過程的結果。盡管這些機制也許是為了調節互惠交換而演化出現的（由此可以想見，這些機制產生了高度依賴于個人所處關系的判斷），但如果后起的社會因素促使人們采取更普遍的友善態度，甚至鼓勵人們發起和維系伙伴關系，無論伙伴有何作為（比如，鼓勵人們把另一邊臉也伸過去挨打），這也不足為奇。做個類比，我們可以設想人類的色覺是為了讓我們區分成熟和不成熟的果實而演化出現的，但這并不意味著區分果實仍然是我們的色覺唯一做的事情。

第二，也許有人反對說，參與到某種互惠關系之中是為了自利，所以這么做的人完全是被自私的目的（盡管也許是“開明的自利”）驅動的，而自利恰恰是道德思維的對立面。這個反對混淆了一些事情。參與到某種互惠關系之中完全可能有助于提高繁殖適應度，但這跟被設計參與這類關系的個體動機如何之間沒有任何關系。連達爾文也犯了這個錯誤。在《人類的由來》（The Descent of Man）一書中，有一段話經常被引來證明達爾文意識到了互惠在人類史前史中的重要性。在這一段話中，達爾文把互惠的起源歸于一種“低級的動機”［（1879）2004：156］。[[75]](#_75_56)喬治·威廉姆斯（George Williams 1966：94）正確地回應道：“我完全不明白這為什么要跟有意識的動機有關系。助他行為要被自然選擇青睞，時不時得到回饋是必須的；但施助者或受助者對此的意識不是必須的。”要補充一下，我也不明白這為什么要跟無意識的動機有關系。互惠伙伴參與到交換關系之中可以是出于自私動機，也可以是出于利他動機，還可以僅僅是由外界環境導致的或者生物體本身內置的反射行為，既說不上不自私也說不上利他。被自私所驅動的互惠生物體的基因，也許完全競爭不過那些被其選中同伴的福利所直接驅動的互惠生物體的基因。而進一步被道德責任驅動的互惠生物體，背叛了對方會有負罪感，它的基因也許還要更有競爭力一些。

至此，本書的第一個任務宣告完成。關于這個話題可說的顯然還有很多，但就“人類做出道德判斷的能力是自然選擇的結果”這個假說而言，本章提供的論據已經足夠讓我們對它青睞有加。本書的第二個任務是回答“那又如何？”這一問題。從這個演化假說出發，可以得出哪些與我們今天的道德判斷、理論和實踐相關的結論？

# 第五章　用演化維護道德

## 5.1　描述的演化倫理學和規范的演化倫理學

本書到目前為止的內容，全都可以被稱作是“描述性的演化倫理學”（“descriptive evolutionary ethics”）：對人類的道德事實上是不是（以及在什么意義上和多大程度上是）自然選擇的產物這個問題進行的研究。我在上文提出的假說是，描述的演化倫理學對這一問題的回答是肯定的。我懇請讀者暫且接受這個假說，即便只是為了接下來討論的需要。從這一假說能得到什么結論呢？尤其是它有沒有倫理學或元倫理學上的涵意呢？

從中我們至少能直接得到一個明顯的評價性結論：道德思維過去對我們的祖先是有用的（至少是在促進繁殖的意義上）。有的人也許想用更明顯的規范性語言加以闡明這種“有用”，宣稱道德以前對我們的祖先“是好的”（或者更嚴格地說，對我們祖先的基因曾經是好的）。我覺得這樣說也不是不可以，但這種“好的”和道德評價差別還很大。這種意義上的“好”，也可以用來說澆水對植物來說是好的，或者熱天氣和干燥的植物對于叢林大火是好的。（對此的進一步討論見下文。）但是，也有許多思想家試圖從描述的演化倫理學中發掘出更明確有力的評價性結論。規范的演化倫理學（normative evolutionary ethics）的支持者們宣稱，描述演化倫理學的正面結論，可以用來辯護一般性的道德或者特定的道德理論、道德判斷和（或）道德實踐。這些人常常被指責為錯誤地嘗試從演化上的“是”（evolutionary“is”）推論出“應當”（“ought”）。我稍后會對這種指責進行評估。另外一種非描述性的演化倫理學還沒有正式的名稱。這種立場的支持者絲毫不覺得描述演化倫理學能用來為道德申辯，他們反而認為描述性的演化倫理學破壞了道德。這一立場并不想從“是”推論出“應該”（所以我不愿將其歸入規范的演化倫理學），相反，我們可以認為，該立場最強的版本宣稱：從演化上的“是”可以推出沒有什么東西是我們在道德上應該做的。這種觀點（我稱之為“用演化論拆穿道德”）是第六章的題目。本章的主題是評估前一種觀點，即規范的演化倫理學。

描述性的演化倫理學還用另一種方式影響著當代人對于道德的思考，即把同倫理決策有重要關系的經驗事實明白揭示出來。倫理決策可以受到方方面面的影響，比如有關某些經濟政策的結果的事實，某種行為導致的有關不幸程度的事實，有關某行為動機的事實，乃至任何在原則上都可以影響倫理決策的事實，而有關人類演化的事實，也同樣可能會影響倫理決策。霍爾沃德·利勒哈默爾說得好：“既然相對而言，我們的道德感受對新的發現來者不拒，那么認為我們的道德感受對關于其自身的發現也有所回應，就再自然不過了。”例如，我們可以想象，關于“什么東西讓我們感到幸福或不幸”的問題，對人類演化的研究結果也許支持了某個假說（例如，參見Grinde 2002）。這將對倫理決策有多重影響；也許這還可以讓我們對某些道德問題改變態度。至于描述性的演化倫理學對道德判斷的這種影響究竟有多么深刻，我在本書的正式立場是：不知道。我的（令人失望的）看法是，要知道答案我們還需要耐心等待。我在這里指出倫理學和人類演化研究的這種關系，只是為了防止讀者把它同更具爭議的那種規范的演化倫理學混為一談，而后者才是本章要討論的題目。

菲利普·基徹爾（Philip Kitcher，早期社會生物學的立場鮮明的反對者）寫道：“人類的倫理實踐是有歷史的，而探究這些歷史的細節完全是合適的。……只要闡述的不是過于簡單，只要不是被過度詮釋（描述性的演化倫理學）是沒有任何問題的。”（1994：440）[[76]](#_76_56)顯然，鑒于本書到目前為止所主張和表示贊同的那些論點，我傾向于認為，在人類倫理實踐的完整歷史記錄里，涉及生物自然選擇的部分將會非常重要。但這里的問題是經驗性的，在找到證據之前，我們不應該盲目相信自己的立場。無論描述性的演化倫理學最后是不是富有成果的研究項目，我們不應該在原則上予以反對。但是，規范的演化倫理學卻名聲不佳。許多人認為，我們無須逐一評估各種形式的規范演化倫理學，因為有一個簡單的先天證明，揭示了任何版本的規范演化倫理學肯定都是錯的。這個證明就是自然主義謬誤。有人說，自然主義謬誤不僅僅推翻了規范的演化倫理學的各個版本，還推翻了任何形式的道德自然主義。

我說的“道德自然主義”，是指以下觀點：道德屬性和關系是存在的，而且還可以毫無齟齬地融入自然主義的世界——即科學可以研究的世界（更詳細的定義見Copp 2004）。功利主義者（道德自然主義者的典型代表）認為，道德就是幸福（也可以用其他類型的利益替換這里說的幸福），而且有關我們在道德上有什么義務（應該）做什么的事實，就是有關什么導致幸福最大化的事實。關于什么東西能產生幸福的事實是因果性的事實和心理事實（都是可以用科學研究的東西），所以按照這種觀點，有關道德價值和義務的事實也同樣是因果性的事實和心理事實。在演化道德自然主義者看來，構成道德價值和義務的基礎的那些在科學上站得住腳的事實，就是有關自然選擇的事實。

我們不能把以上這兩種自然主義觀點和全局自然主義（global naturalism）混為一談。后者是更具一般性的立場，它認為只有那些能夠被納入統一的科學框架內的事物，我們才有理由相信其的存在。全局自然主義者希望在以下意義上“把道德自然化”：給道德體制和實踐，道德心理學，以及道德譜系學提出一套科學上站得住腳的解釋。但是，持此觀點的人還不能算作道德自然主義者，因為持此觀點和認為“所有道德信念都是錯誤的”或“道德語言壓根就不是用來指稱屬性和關系的”并無矛盾。為了更好地體會“把道德自然化”這個短語的模糊之處，我們可以類比一下“把巫婆自然化”。任何有科學頭腦的人都會想要在以下意義上“把巫婆自然化”：給有關巫婆的信念和實踐活動，提出一套有經驗證據的可靠解釋，一套有關這種文化現象的社會學、心理學和人類學。但是這項可敬的探索到最后不必（也不應該）為那套圍繞巫婆的話語提出辯護，也就是說，不必證明自然主義的世界觀可以接受巫婆的存在。在后面這種意義上把巫婆自然化可就完全是另一碼事了（也就是誤入歧途了）。在本書中，“道德自然主義”（以及衍生的短語“把道德自然化”）指的是后一種立場，即認為道德屬性確實存在。一個全局自然主義者可以否認這種立場；事實上，也許恰好是對全局自然主義的信奉讓人拒斥道德自然主義，轉而接受道德虛無主義（就好像同樣的全局自然主義讓我們對巫婆的存在持虛無主義的態度。）自然主義謬誤的靶子不是全局自然主義，但在有的人看來，它不僅會對道德自然主義，因此也會對演化的道德自然主義造成毀滅性的打擊。

但需要注意的是，“規范的演化倫理學”比“演化的道德自然主義”的范疇更廣。規范的演化倫理學家一心想要借助有關演化的事實來為道德進行某種“辯護”，但證明道德屬性確實存在且基于某些有關自然選擇的事實，這只是實現此種辯護的一種方式而已。在本章后面的部分，我將考察其他的方式。（這些方式形態各異，這就是我為什么讓“辯護”這個詞保持模糊和非正式。）但是，讓我們先聚焦于演化的道德自然主義者，看看對他們來說自然主義謬誤究竟是不是致命一擊。

## 5.2　自然主義謬誤

從自然的“是”得出倫理的“應當”的嘗試，以往被稱作“自然主義謬誤”，這一稱號算是恰如其分。

——保羅·R.艾利希（Paul R.Ehrlich 2000：309）

從演化（乃至從任何觀察到的自然現象）中提取出價值的念頭，已經被G.E.摩爾（G.E.Moore）徹底地打破了。他把這種錯誤稱為“自然主義謬誤”……摩爾并不是第一個質疑從“是”到“應當”的推理的人。

——羅伯特·賴特（Robert Wright 1995：30）

當代哲學的一種陳詞濫調是我們不能從“是”推出“應當”。進行這種推論，就是犯了“自然主義謬誤”。這個術語來自G.E.摩爾的經典著作《倫理學原理》（1903）[[77]](#_77_56)。

——丹尼爾·丹尼特（Daniel Dennett）

我們可以毫不費力地找到下面這樣的說法：G.E.摩爾在1903年指出，“應當”不能從“是”中推論出來，他還指責任何這樣做的人都犯了“自然主義謬誤”。我從書架上拿出幾本書，隨便掃一遍就找到了以上三段引文。然而，這些說法都錯得讓人大惑不解。摩爾確實把某些問題稱作“自然主義謬誤”，但肯定不是所謂的從“是”推出“應當”的錯誤。對此有懷疑的人，我請你們翻開摩爾的《倫理學原理》［Principia Ethica（1903）1948[[78]](#_78_54)］，找出任何提到這種推論的地方，或是提到“應該”和“是”，或“事實/價值之間的鴻溝”的地方。摩爾設想的“自然主義謬誤”指的完全是另一回事。在本節中，我將陳述（并批判）真正的自然主義謬誤，及其可疑的同伙，即“開放問題論證”。這將帶我們稍稍偏離主題，進入哲學史的領域，但我覺得要是能幫助澄清相關的問題，這樣做還是有道理的。下一節我將討論并評估那種被誤稱為“自然主義謬誤”的論點：我們不可能有效地從“是”推論出“應當”。

在摩爾看來，任何想要給“善”下定義的嘗試都犯了自然主義謬誤，因而注定要失敗。“給善下定義”（“def ine good”）這個短語在今天的人們聽起來頗成問題，但這確實是摩爾偏愛的表達方式。他非常明確地指出，他對定義“善”這個詞的含義沒有興趣：“我想關注的只是在我看來（無論是對是錯）這個詞一般而言所表示的東西或理念。我想發現的是那個東西或理念的性質，而在這個問題上，我是非常想要能得到同意的。”因為摩爾感興趣的是分析“善”這個詞指稱的對象（即善的屬性），所以當他說“給善下定義”的時候，我們要理解他追求的是以往人們說的“實質定義”，而不是名義定義。這個區分可以上溯到蘇格拉底。名義定義解釋的是一個語詞在語言中是怎樣使用的，而實質定義解釋的是一個東西是什么，最好是揭示其本質特征。（我的意思不是說這個區分是沒問題的，甚至也不是說它是有用的；我只是在解釋摩爾的有些說法。對這個區分，Robinson 1950一文中有些有益的探討。）

必須要記住的是，在摩爾自己眼中，自然主義的謬誤不僅體現在定義善的嘗試中，也體現在定義快樂和黃色的嘗試中。這表明善的“評價性”（即其“應當性”）跟這種謬誤中沒有太大關系。事實上，摩爾對該謬誤最簡潔的表述針對的是黃顏色：

這其實是個非常簡單的謬誤。我們說某個橙子是黃色的，但這樣說并沒有迫使我們承認“橙子”的意思就是“黃色”而已，或者只有橙子才是黃色的。[[79]](#_79_54)要是橙子同時也是甜的，難道這迫使我們承認“甜的”和“黃色的”意思是完全一樣，或必須被定義為“黃色”嗎？……除非黃色的確除了“黃色”再沒有其他任何的意思（除非黃色完全不可定義），否則說橙子是黃色的毫無意義。（14）

摩爾聲稱，有的東西是不可定義的，所以想要定義這樣的東西當然也就是有問題的。這些東西包括評價性的性質，比如道德的善，但也包括根本是非評價性的性質，比如黃色。摩爾認為，這些東西無法定義，是因為它們是“簡單的”而且“沒有組成部分”（9）。與此相反，他認為馬（我覺得更好地說法是馬性（horseness））是可以定義的，因為馬是個復雜的物體。他大概會很喜歡下面這段話，出自狄更斯的《艱難時事》：

“比策，”托馬斯·格雷格林說，“來說說你對馬的定義。”

“四足動物。食草。有四十顆牙齒，包括二十四顆臼齒，四顆犬齒，十二顆門齒。春天換毛；在沼澤地區也要換蹄。馬蹄很硬，但還是需要上蹄鐵。可以通過牙口知道馬的年齡。”

“二十號女孩，”格雷格林說，“你這下知道馬是什么了吧。”

摩爾認為，善沒有這種分解式的定義。他在論證的過程中，從來沒有把善的評價性提出來作為其不可定義的理由，也沒有訴諸“應該”和“是”之間的所謂邏輯鴻溝。事實上，《倫理學原理》后面的部分以及摩爾其他的論著都顯示，摩爾相信有關內在價值的陳述語句是事實性的（參見Bruening 1971）。他說按照“‘純粹的事實陳述’這個短語的一種用法，‘p是純粹的事實陳述’。并不等于‘p不是規范性的’”。“很明顯，”他繼續道，“像‘我應該去做X’這樣的陳述句是規范性的，但它是否就不能還原成事實陳述句或存在陳述句，這就不怎么明顯了。”（Moore，收入Schlipp 1942：568）摩爾不但完全沒有論證價值陳述和事實陳述是兩碼事，他還對兩者間存在的還原關系保持樂觀！他所堅持的是，形如“X是善的”或“善是Y”這樣的句子，不應該被當作定義（指前面探討過的分解定義）。換言之，只要句子里的“是”是用來連接謂語，而不是用來表示等同的，那么說“X是善的”或“善是Y”就還是可以的。從根本上而言，這才是摩爾的自然主義謬誤：把用謂詞修飾誤當成是下定義（在摩爾那里，后者指的是分解式的定義句），就是大錯特錯。說善如何如何，沒有問題，但說善是什么，這就不行，因為善事“簡單的，不可定義，不可分析的”（21）。

所以，重要的問題在于：我們為什么要贊同摩爾的說法，認為善不可定義？這就輪到他的開放問題論證（Open Question Argument，OQA）出場了。在過去的一個世紀里，人們對開放問題論證（OQA）的內容和步驟，就像對自然主義謬誤一樣，也形成了某種“傳說”；對這個論證的流行解讀跟摩爾本人的表述似乎很有些奇怪的偏差。流行的版本如下：

假設道德自然主義者（規范演化倫理學的支持者是其中一種類型）提出了這種一般形式的分解式實質定義：

自然主義：對任何x而言，x是善的當且僅當x是N

N指的是某些自然屬性。無論你設想的N是什么，都可能有某個完全掌握了這套語言的人，確信某個東西是N，但并不確定它是不是善的（即“它是善的嗎？”的問題依然是懸而未決的）。這表明原來的自然主義肯定是錯的。

這是個有缺陷的論證。自然主義中的“當且僅當”可能有兩個意思，我們必須要區分開來。它可以指一種語言上的、先驗（a priori）的關系（例如，“對任何x而言，x是個單身漢當且僅當x是個未婚男子。”里的“當且僅當”），也可以指一種形而上的、后驗的關系（例如，“對任何x而言，x是水當且僅當x是H2O。”里的“當且僅當”）。邁克爾·史密斯將其分別稱為“定義的自然主義”（“def initional naturalism”）和“形而上的自然主義”（“metaphysical naturalism”）（1994：27ff）。如果自然主義者提出的是一套形而上學理論，那么顯然OQA毫無效力。如果某人不知道相關的分子理論，他可能確信杯子里的液體是水，但不確定（甚至干脆否認）杯子里的是H2O。但這種可能性完全不能讓人們對“水是H2O”的真實性產生懷疑。事實上，按照正統的觀點，“水是H2O”是必然真理。

人們也許可以更寬容一些，把摩爾理解成只是在攻擊定義的自然主義。但這還是有嚴重的問題。首先，一整類的自然主義（形而上學的自然主義）都不在攻擊范圍內；其次，這種限定跟摩爾的原文有沖突。在可以允許的定義中，他最喜歡的例子（馬的定義[[80]](#_80_52)）是用其肝臟、心臟等，以及這些東西之間的關系來定義馬的，而這些看起來都是有關馬的后驗真理。第三，即使攻擊范圍只限定在定義的自然主義，OQA能否成功仍然比較可疑。簡單地說，在我們眼里先驗真理不夠明顯，因而需要一番智力上的耕耘才能確定其真實性（參見Smith 1994第二章）。例如，假設對知識的概念的正確分析是得到證成的真信念。一個掌握了語言的人，可以既確信某個事物是知識，同時又懷疑它是不是得到證成的真信念。即便如此，“對任何x而言，x是知識當且僅當x是得到證成的真信念”依然可能是正確的，而且還可能是先驗必然的真理。

對這個可以輕易駁斥的傳說版本，我們的討論就到此為止。但是仔細研讀摩爾的原文，我們可以發現，OQA其實是相當復雜的論證，雖然也沒有比傳說的版本更成功。摩爾首先引入了“N”的一個特定的、據說是任意的例子（其實來自于伯特蘭·羅素）：

自然主義\*對任何x而言，x是善的當且僅當x是我們渴望去渴望（desire to desire）的東西。如果OQA確實像上面傳說的那樣，可以想見，隨后出現的問題應該是：

（1）我們渴望去渴望X嗎？

以及

（2）X是善的嗎？

再接下來，該論證就會指出，可能某個掌握了相關語言的人確信（1）的答案是肯定的，但同時不確定（2）的答案是什么，而這就表明自然主義\*是錯誤的。然而，摩爾不是這樣論述的。他說的其實是，就像可以問X是不是善的，我們也可以“有意義地”問道，“那個如此定義的復合體本身是不是善的”（15）。換言之，我們可以問道

（3）渴望去渴望X是善的嗎？

根據自然主義\*，這就等于問

（4）我們渴望去渴望去渴望去渴望X嗎？

這樣一來，摩爾認為，自然主義\*就被駁倒了，因為在問（3）的時候，“我們心里想的不是像（4）這么復雜的東西”。他認為，也許“如果某物是善的，我們就會渴望去渴望它”，乃至“如果我們渴望去渴望某物，那么它就是善的”，這兩句話都是正確的，但“我們能夠對這些命題產生懷疑，這個事實本身就清楚地表明我們心里有兩種不同的概念”（16）。

這個論證確實揭示了“……是善的”這個謂詞有些不尋常。我們可以說“X是善的”，也同樣可以說（至少在語法上是正確的）“X是善的，這是善的”（“It is good that X is good.”）。很多謂詞就不是這樣：我們可以說“X是方的”，但是說“X是方的，這是方的”（“It is squire that X is squre.”）就是莫名其妙了。換言之，正如我們在第二章見過的，“……是善的”這個謂詞可以合理地應用于行為和事態（states of affairs），但“……是方的”只能合理地應用于物體，不能用于事態。這似乎突出了自然主義要滿足的一種條件：無論N是什么，它最好也具有這種特征。然而，對于這個論證，能說的好話大概也就僅此而已了。

首先，我們也許會懷疑，摩爾所謂任意選擇的那個關于善的自然主義理論（善是“我們渴望去渴望的東西”）是不是真的沒有一點責任，因為正是它導致了啰唆而且有些愚蠢的（4）。也許我們“心里想的”確實不是像（4）這么“復雜”的東西，但（4）如此復雜正是緣于摩爾精心選擇的例子！更重要的問題是，我們“心里想的”是什么，其實并不相干。如果對方提出的是形而上的自然主義，那內省就不是對同一關系的檢測手段。不了解化學的人，也許可以合理地懷疑水是不是H2O，但“水必然是H2O”仍然可以是正確的。歸根結底，OQA無非是摩爾想讓他的讀者直接“看到”善不能等同于任何其他屬性，除此之外沒有什么更具說服力的東西；這表明他在元倫理學的層面上是個直覺主義者（Fanaei 2003）。

另外值得注意的是，摩爾所表述的整個論證都是關于我們能夠問什么。他認為，如果自然主義\*是正確的，那么一個人問（3）的時候就是在問（4）。但這是錯的，因為提問表達的是隱晦語境。如果露易絲·蓮恩問道，克拉克·肯特是不是（Clark Kent）拯救了世界，她并不是在問是不是超人拯救了世界，雖然超人和克拉克·肯特是同一個人。事實上，對于摩爾最喜歡的、成功的分解定義的例子，顯然有人也可以問“X是匹馬嗎？”，但他并不因此就是在問“X是否有四條腿，一個頭，一顆心臟，一個肝臟等，而且這些東西彼此之間都有一定的關系？”同理，問（3）的時候不一定是在問（4），但這沒有排除自然主義\*為真的可能（就算我們將其理解為定義的理論）。

現在做個小結：摩爾的自然主義謬誤同禁止從“是”推出“應該”沒有多大的關系；該謬誤其實是要說，我們不能定義無法定義的東西，因此就不應該做出這樣的企圖。作為回應，我們首先要注意，道德自然主義者（因此演化道德自然主義者）可能不怎么在意這種禁令，因為他的目的遠遠不像是要給道德上的善提出一個分解式的定義。只要他不是想那樣做，那么他沒有必要在不經意間挑戰摩爾的主張。但是就算道德自然主義者想要給出一個那樣的定義，摩爾用來支持他看法（即善是“簡單的，不可定義的，無法分析的”）的唯一論證就是開放問題論證。我們已經看到了，這個論證的問題很嚴重，它所依靠的是困擾著20世紀前期哲學的一些有關必然性、先驗性和分析性的混亂觀點，直到20世紀中期，通過諸如斯瑪特（J.J.C.Smart）和普萊斯（U.T.Place）[[81]](#_81_52)等哲學家的工作，這些混亂才得到厘清。道德自然主義者（尤其是演化道德自然主義者）不會被這個論證觸動，而這種反應也是恰當的。

“自然主義謬誤”這個術語，偶爾也被用來形容另一種觀點：從“X是自然的”得出“X是善的”（我也見過有的說法是從“是什么”得出“應該怎么樣”），這就犯了這種謬誤。這同樣跟摩爾說的謬誤沒什么相似之處，雖然他無疑會同意這確實是個嚴重的錯誤。但沒有什么嚴肅的思想家會認為（某種形式的）自然和道德上的善之間有簡單直接的聯系，雖然偶爾會有學術圈外的人這么說。無論在大眾眼中能有多少吸引力，這顯然是錯誤并愚蠢的說法。但需要指出的是，光是錯誤和愚蠢還擔當不起“謬誤”的榮譽。謬誤是推理中的錯誤，而不是錯誤的信念。如果有人覺得，同性戀是非自然的，而且所有非自然的事情都是錯誤的，因而同性戀是錯誤的。他的推理毫無瑕疵，沒有犯任何謬誤；有問題的只是他的論證前提所表達的錯誤信念。給哲學對手的觀點貼上“謬誤”的標簽，常常只是卑鄙的修辭把戲。我想稱之為“謬誤的謬誤”（“the fallacy fallacy”），但這樣稱呼就落入這個謬誤了。

## 5.3　從“是”推出“應當”

馬可·波羅從一些阿拉伯海員那里聽說了“Madagascar”[[82]](#_82_52)（更準確地說是“Maydeygastar”）這個名字；他自己從來沒有接近過那里。但在那次交流中，波羅不知怎的認為這個名字指的是非洲近海的一個大島，盡管實際上它指的是非洲大陸的一部分。這導致了歐洲的地圖繪制師長期的混淆。然而，要是有人還堅稱現在把那個島叫作“Madagascar”就是犯了錯誤，可就太迂闊了。同樣地，已經有太多的人宣稱，自然主義謬誤說的就是不能從“是”推出“應當”，我現在幾乎都愿意承認這確實是這個詞組的意思，因為倘若有足夠多的人犯某個語言錯誤，那就不再是錯誤了。對這個詞組的混用在哲學界已經根深蒂固，難以改變，就像“Madagascar”已經難以恢復本義一樣。無論如何，就像摩爾認為他的自然主義謬誤摧毀了所有形式的道德自然主義，許多人認為，用“不能從‘是’推出‘應當’”這個一般性的先驗原則，也可以有一樣的打擊效果。兩者是不同的主張，但都值得我們的重視。所以，我們能否從“是”推出“應當”呢？

如果這個問題的意思是“是否存在一個個在邏輯上有效地論證，其前提不包含‘應當’，但其結論包含有這個詞？”，那答案無疑是肯定的。比如以下論證（出自Prior 1960）：

P1：巴黎是法國的首都。

因此，要么巴黎是法國的首都，要么你不應該偷香蕉。

或者這個：

P1：要么巴黎是法國的首都，要么巴黎不是法國的首都。

因此，你不應該偷香蕉。

這兩個都是怪異的論證，但也都是完全有效的。前者運用了經典邏輯里常見的析取引入規則，后者則利用了經典邏輯的一條定理：從矛盾的前提可以推出任何結論。［先設定前提（i）P且非P，用簡化規則可以由此推出兩個命題：（ii）P和（iii）非P。對（ii）使用析取引入規則，對于任意的Q，都可以得到（iv）P或Q。對前提（iii）和（iv）使用析取消除規則，則可以得到結論Q。］

但可能有的人會懷疑這些邏輯把戲。查爾斯·皮登（Charles Pigde）抱怨道，這些例子的結論并不是真正地（genuinely）從前提推出來的，因為即便把包含道德謂詞的結論（“你不應該偷香蕉”）替換成語法上相同的命題（“珍妮正坐在鮑勃的旁邊”），論證還是一樣有效。在這兩種情況里，結論“相對于前提而言是空洞無物的”（1989：133）。假如“不能從‘是’推出‘應當’”的意思果真如此（即其意為不能從“是”非空地推出“應當”），那以上那些反例就沒有效力了。

但是重要的問題是“這有什么要緊呢？”在經典邏輯（加上對于有效性的正統的、塔斯基式的理解）中，正如包含“應該”的非空結論，不能從不包含該謂詞的前提中推論出來，同樣地，明確關于刺猬的非空結論，也不能從不包含“刺猬”一詞的前提中出來（Pidgen 1989）。或者不妨考慮如下論證：

P1：我的杯子里有H2O液體。

P2：我正要喝我杯子里的液體

因此：我正要喝水。

按照塔斯基式的標準，這是個無效論證。如果將其全部翻譯成謂詞演算的語言，我們不難找到一個使P1和P2為真但結論為假的模型。[[83]](#_83_52)即便如此，我們對于水和H2O是同一種物質（事實上人們常常認為二者必然是同一的）的信心也不會受到什么打擊。道德自然主義者（因此演化道德自然主義者）可以堅稱道德屬性等同于（或隨附于）自然屬性，但同時否認在有關自然世界的命題和涉及道德價值和義務的命題之間存在著語義的或演繹的的關系。類似地，我們一般都默認生物學事實符合由物理學所刻畫的世界觀，雖然從有關物理學上的物體和法則的命題前提，到生物學上的命題結論，似乎并不存在有效地演繹。我們也無法從只提及了原子及其行為的前提，推出有關桌子椅子的結論，但我們都同意桌子椅子符合物理學描述的世界觀。當然，也許自然主義者不能夠讓我們信服道德價值和家具都以同樣的方式符合自然主義的世界，但重要的是，正如威廉·弗蘭肯納認識到的那樣（1939），這是要靠哲學論證解決的問題，而不是一開始就武斷設定的東西。

讓我們先來假定，演化的道德自然主義者認為，有關人類演化的某些屬性（姑且稱其為“E”）和“道德應該”是同一種屬性。現在考慮一下論證：

P1：約翰的行為ϕ具有屬性E。

因此：約翰在道德上應該做行為ϕ。

這顯然是個無效論證，但自然主義者無須否認其無效性［羅伯特·理查茲（Robert Richards）確實不認為這是無效的，后文將討論他的觀點。］演化自然主義者只是認為，該論證是省略三段論（也就是說，它需要另一個前提才能成為有效）：

P2：對于任何行為x而言，x具有E當且僅當x是道德上應該做的。

誠然，P2幾乎肯定是富有爭議的，也不是能讓人即刻贊同的。然而這并不是問題所在。自然主義者認為P2是真的，但他并不認為P2之為真是顯而易見和無可爭議的。也許有人會進一步抗議道，當道德自然主義者試圖說服我們接受P2時，他所做的無非還是讓我們相信從P1可以推出那個結論，但既然我們之前已經拒絕接受了這個推論，那現在怎么會被說服呢？可是這樣說就錯了。對原來論證的駁斥，關注的是其無效性，但自然主義者在擁護P2時，他并不是在說E—命題和“應該”—命題之間有演繹關系，而只是在說P2表達了真理。只是因為P2連接了“應該”和“是”，就一口咬定P2不可能是真的，這種反駁無非是回避了問題而已。毫無疑問，要讓我們接受P2，自然主義者要花費一番功夫，但這里的重點在于，光是揭露某種嚴重的無效性，或者光是因為犯了某些謬誤，還不足以使自然主義的主張一敗涂地。

概述一下之前的內容：規范的演化倫理學并不局限于演化的道德自然主義。規范的演化倫理學試圖用某種方式借助演化來“維護”道德，但（我們稍后將看到）確立某種形式的道德自然主義（比如揭示道德屬性等同于或隨附于某些自然屬性）并不是唯一可行的維護策略。所以，哪怕摩爾的自然主義謬誤和是/應當之間的鴻溝成功摧毀了道德自然主義，規范的演化倫理學家還是可以去探索別的途徑。然而，我們已經看到了，這些批評的考量遠沒有成功。兩者都犯了轉移主要問題的錯誤，早在20世紀中期就應當被淘汰了。那么歸根結底，演化的道德自然主義和（更具一般性的）規范演化倫理學是不是最終還是可行的立場呢？我覺得不是。在第六章中，我將提出一些論證，反對一般的道德自然主義（這些論證要是能成功的話，也一并駁斥了任何演化版本的道德自然主義）。至于演化的道德倫理學這個更加寬泛的范疇，則沒有一擊致命的一般論證，因為相關提議的種類太多，我們只能一一考察，同時也留意找出那些論證里反復出現的弱點。這就是本章剩下的部分要進行的工作。（接下來要討論的這些規范演化倫理學家，有的意圖提出一種道德自然主義，有些不，還有的我并不確定。）重點要指出的兩個反復出現的問題是：（1）對于道德判斷的認知成分，規范的演化倫理學家經常表現出一種明顯或不明顯的（似是而非的）忽略，以及（2）規范的演化倫理學家常常找出演化假說所蘊含的某些非道德的規范性，然后錯誤地將其歸為道德價值。

## 5.4　從演化的角度維護道德：理查茲

1986年，在《生物學和哲學》雜志的第一期上，羅伯特·理查茲（Robert Richards）為“修正的”（revised）演化倫理學［被“修正”的是過時的版本，比如赫伯特·斯賓塞（Herbert Spencer）的理論；這些理論我在此不做討論］進行了詳細的辯護。理查茲首先想要說明，一般而言“應當”可以怎樣從“是”推出來；然后他做出了具體的嘗試：用相關的演化前提來實現這種推演。我將依次討論這兩個論證。

我們之前看到了，道德自然主義者不需要為自己辯護說在經驗的/描述性的命題和道德命題之間有演繹的關系，但如果這種關系能得到支持，對于道德自然主義而言也是好事一樁（也就是說，建立這種聯系對于道德自然主義是充分但非必要的。）理查茲想要做的，就是借助任何邏輯論證都必須有的推理規則來為這種關系辯護。

考慮一下劉易斯·卡羅爾設想的阿基里斯和烏龜之間的辯論。阿基里斯想要讓烏龜承認以下論證是有效的：

（A）跟同一個物體相等的兩個東西，彼此也是相等的。

（B）這個三角形的兩條邊，是同一個物體相等的兩個東西。

（Z）這個三角形的兩條邊彼此相等。

烏龜承認A和B，但拒絕承認Z，因為他拒絕承認以下命題

（C）如果A和B是真的，則Z必須是真的。

于是阿基里斯把C加入到論證的前提里，然后要求烏龜承認現在的前提蘊涵Z，這是因為，他說道：“如果你承認A、B和C，那你就必須承認Z。”烏龜反駁道，

“那為什么我必須承認它呢？”

“因為前提在邏輯上蘊涵著它。如果A、B和C都是真的，那么Z必須是真的。我想你不會對此也有爭議吧？”

“如果A、B和C都是真的，那么Z必須是真的。”烏龜若有所思地重復道。“這又是一個假定，對吧？要是我不覺得它是真的，我就可以承認了A、B和C，但仍然不承認Z，不可以嗎？”

“你這么想是可以的，”坦率的英雄直言道，“雖然這就顯然非常愚鈍了。但這依然是可能的。所以我得讓你再承認一個假定。”

“很好。只要你把它寫下來，我很愿意承認它。我們可以稱之為

（D）如果A、B和C都是真的，則Z必須是真的。”

“你記到你的筆記本上來了嗎？”

“記下來了！”阿基里斯高興地喊道，一邊把鉛筆放進套里。“我們終于結束了這場理念競賽！現在你承認了A、B、C和D，那你當然也會承認Z。”

“是嗎？”烏龜無辜地說道，“讓我們把話說得很清楚。我是承認了A、B、C和D。我要是還不承認Z呢？”

“那么邏輯會迫使你承認的！”阿基里斯勝利般地答道。“邏輯會告訴你‘你不能一廂情愿。既然你現在承認了A、B、C和D，那你必須承認Z！’知道了嗎，你別無選擇。”

“邏輯教導我的好東西都值得寫下來，”烏龜說道。“所以請寫到你的筆記本上吧。我們將稱之為

（E）如果A、B、C和D都是真的，則Z必須是真的。

要是我不接受這個前提，顯然我就不需要接受Z。所以這一步是必須的，你明白了嗎？”

“我明白了。”阿基里斯說道；語氣中透露著一絲沮喪。

可憐的阿基里斯。顯然，他本應該說諸如“如果P和P則Q都成立，那么Q成立”（假言推理）并不是額外的前提，而是導致從前提轉入結論的推理規則。毫無疑問，烏龜會答道，“這些推理規則是怎么來的？”這是個很好的問題，也正是卡羅爾想讓我們思考的問題，他還給我們表明了一種不該給出的答案。理查茲的答案則是，我們是從“理性的人們”的“信念和實踐活動中”得出的這些規則（284—285）。隨后他宣稱，類似地，我們也可以從有道德的人們的信念和實踐活動中，得出連接經驗事實和道德價值的命題，這些命題的功能不是作為論證的前提，而是作為推理規則，批準從前提得出結論，也因此允許從“是”得出“應當”。

理查茲的答案要是能成功，那當然很不錯。但是否真有這樣的推理規則，實在是極度可疑。至少有三個原因讓他的論證缺乏說服力。首先，為什么我們就只能借助在理性的人們看來是自明的東西，否則我們就不能更好地證成假言推理規則，這顯然說不清楚。而且我們怎么確定誰是“理性的”呢？條件之一不就是看他們是否肯定和遵循假言推理嗎？其次，如果我們要證明從描述性的前提可以推出道德結論，為什么就應該要借鑒有道德的人們的信念和實踐活動，這也不清楚。為什么不借鑒理性的人們的信念和實踐活動呢？把目光局限于有道德的人（無論是誰），就破壞了整個工作，因為在傳統上，從“是”推出“應當”的挑戰就在于如何讓理智的道德懷疑論者相信，她所合理接受的經驗事實意味著她也接受了某些道德結論。讓一群功利主義者同意某個行為因為能使幸福最大化因而在道德上是有義務去做的，或者讓一群康德主義者同意，就某個行為而言，要是其準則的普遍化形式不能被意愿成為一條自然法則，該行為就是被禁止的——這些都不是難事。但難題在于讓還沒有同意道德觀點的人同意。第三，一個東西被普遍接受，我們還不能由此就得出結論說它是個某種意義上的推理規則。幾乎人人都同意“老虎是四足動物”這個命題，但這充其量意味著你要是把“老虎是四足動物”放在論證里，作為一個前提，那么幾乎沒有人會反對。這不等于說是該命題是個演繹規則（就類似于假言推理），可以讓你從只提及老虎的前提得出提及四足動物的結論。因此，即便有些連接經驗事實和道德價值的命題，人人都同意，這并不必然表明有一條特殊的道德推理規則；也許只是意味著有一個人人都同意的（包含“應該”的）前提，在這種情況下，我們還是沒有發現光從經驗前提就能得出的“應該”。

歸根結底，理查茲的工作在于指出，我們在給任何東西尋找理據的過程中，總是要安于在某些終點（這些終點在我們看來是直覺上是顯而易見的），否則人們總是可以追問“這是怎么證成的呢？……哦，好的，但那又是怎么證成的呢？”，而我們就永遠無所歸附了。這當然是有道理的。但問題依然存在：怎樣才算是適當的證成終結點呢？光是指出不管是哪總得有一個終點，然后選一個你最中意的競爭者（這個終點剛好也能支持你的理論），這還是不夠的。

關于理查茲從“是”推出“應當”的嘗試，就先討論到這里。然而，正如我說過的，他的演化道德自然主義并不應該要求在描述命題和規范命題之間建立演繹的關系。現在我們再來看看他在演化的基礎上維護道德的實質嘗試。他的論證的核心部分出現在以下的段落：

……證據表明，事實上演化造成了人類為了社群的利益而行動；但我們說的有道德的意思就是為了社群的利益而行動。因此，由于人類是有道德的生物（這是演化造成的不可避免的情況），每個人都應當為了社群的利益而行動。（289）

為了論證的方便起見，我們先姑且承認第一句話。但第二句話（我們的“道德”這個詞的意思就是為了社群的利益而行動）是錯誤的。也許人類的道德感是為了增進社會凝聚力而演化出來的，但不能由此推論出“道德”這個詞的意思。（同理，也許人類的雙足屬性是為了讓我們大草原上的祖先留意捕食者而演化出來的，但不能由此推論出“站立”這個詞的意思就是留意捕食者。）另外，不利于社群的道德行為，以及有利于社群的非道德行為，例子都很容易找到。后者的例子：設想我是成吉思汗的黨羽，而且我的自然本能是為了我的社群利益而行動。設想我和我那些皮膚黝黑的同伴們決定，要實現這一目的，最好的方法是把垃圾倒入我們遇見的所有其他社群那里。這是道德的嗎？也許我和我那些粗魯的伙伴覺得這是道德的（在我們看來這是道德的），或者我們嘲弄道德規則這一觀念本身。（“道德，狗屁道德！”我們大笑著，又把一顆頭顱釘到長釘上。）為了社群的利益而行動常常是道德的，就像交朋友和講真話一樣，但我們能輕易想到這些行為類型的丑惡個例，這本身就說明這樣的行為都不是我們說的“有道德”的意思。

然而，就算只是為了論證的方便，假設確如理查茲所說，為了社群的利益而行動和道德的行為是一回事，以及演化把我們設計成要為社群的利益而行動。那么，由此可以輕易推論出，演化把我們設計成為道德的。但我們該怎么理解他的結論“每個人都應該為了集體的利益而行動”呢？如果人類被構造成要為社群利益而行動（即被構造成有道德），是“演化造成的不可避免的情況”，那么我們原本預期的結論是每個人都將會為了社群的利益而行動（也就是說將會有道德）。那么結論里的“應當”突然出現是怎么回事？我唯一說得通的解釋是，這里預測的“應當”（正如在“雨應當一會就要停了”里的“應該”）；實際上理查茲也承認這一點，將結論里的“應當”和“剛才閃電了，現在應當有雷聲了”里的“應當”相提并論。但這種用法就非常不合適了，因為我們知道，道德的“應當”和預測的“應當”有天壤之別。一個偵探談起連環殺手，可能會說“他應當一周之內還會有行動”，意思是在一周結束前他很可能再次行動，但這跟說殺手在道德上應當再次進行謀殺完全是兩回事。反過來，人們可以說殺手在道德上應當自首，但并不因此暗示這種情況很有可能發生。

規范的演化論理學家往往太過執著于從“是”推出“應當”，以至忘記了他們想發現的是哪種類型的“應當”。理查茲就是個典型的例子，他得到預測的“應當”就心滿意足了。那要是這樣就足夠了，那么從演化論（或者任何經驗事實）中找到道德規范就確實是件很容易的事情。物理學家可以說某個粒子應當偏離運室里的離子束，動物學家可以說鵝群應當不久后就飛往南方，諸如此類。一旦我們注意到，同一個詞既可以表達預測關系也可以在道德上規定行為，這只是英文的純粹巧合，其他語言并不一定如此，我們就應該對這種從“是”推出“應當”的方式心生懷疑。

## 5.5　從演化的角度維護道德：坎貝爾

理查茲的文章發表十年以后，在同一份期刊中，里奇蒙德·坎貝爾（Richmond Campbell 1996）試圖在人類道德是先天的這一經驗假說的基礎上，對道德進行更有限度的維護。他論證的不是道德自然主義本身；他自己限定的目標，僅僅是確立上述經驗假說可以表明道德是得到了證成的。他小心翼翼地指出，他只是給人類有某種（而不是完全沒有）道德提出理據，而不是給某個有特定內容的道德體系提供理據。他的論證相當簡單，而且如果加以一定的條件，也可能是成功的。但是清楚地理解了這些條件，亦將揭示其成功的限度。下面是該文極重要的段落：

……這個論證依賴于以下規范的但非道德的原則：如果具有某種道德，比起道德完全缺失，能極大地改善群體里每個成員的生活前景，那么，對于群體里的每個成員而言，具有某種道德，相對于道德完全缺失，是得到證成的。［把這個原則稱為“對大家都極有利”（“overwhelmingly mutually advantageous”,OMA）原則。］對道德的存在的生物學解釋，意味著，相對于道德完全缺失，具有某種道德能極大地改善群體里每個成員的生活前景，由此我們可以推論出（借助上述原則）具有某種道德，比起道德完全缺失，是得到證成的。（Campbell 1996：24）

我們可以率先對“道德的演化解釋證明了具有某種道德，比起道德完全缺失，對于每個人來說都要更好”這個假設發難。本章開頭就指出了，演化假說并不意味著道德現在是有用的。它意味著（A）道德過去是有用的，以及（B）道德過去對于我們祖先的基因是有用的。對于（B），坎貝爾意識到，他需要明確肯定至少在相關的討論范圍內，我們可以相信對某人的繁殖適應性有利的，一般也對其本人有利。這倒是不無道理，但是論證想要完整詳盡的話，這一說法還需更多的證據。坎貝爾似乎對于（A）也有所意識，因為在討論的時候他用的都是過去式：“……因為在道德演化形成的時期，這兩種因素（個人幸福和基因繁殖適應性）是高度相關的。”（25）但是他似乎沒有注意到，證明道德過去對每個人都有用（即證明，根據OMA，道德過去是證成的），并不足以證明道德現在還有用（即證明，根據OMA，道德現在也是證成的）。出于同樣的理由，證明吃掉所能找到的全部甜食的強烈傾向在過去對我們的祖先有用處，因此曾經是（在某種意義上）證成的，這還不足以證明今天這種傾向依然是有用的或證成的。另外，倘若坎貝爾著手填補論證的這一步（也就是說，倘若他展示更多的經驗證據，來支持以下結論：現在我們具有某種道德，比起道德完全缺失，能極大地提高我們群體每個成員的生活前景）。那么至于人類史前時期的情況就完全不相干了。換言之，坎貝爾面臨著兩難的處境：要么他是從演化假說得出結論，但是他的結論就只能是有關過去對于我們祖先來說，什么是證成的；要么他就要試著證明道德現在對我們有用，但那樣的話演化假說就顯得贅余了。

然而，更重要的問題在于坎貝爾是否采用了正確的證成概念。可以比較這個例子：假設我們可以證明，有宗教信念（比起完全沒有）改善了群體里每個人的生活前景。也許相信自己的生活是某種宏偉藍圖的一部分，或者相信冥冥中有著更大的力量（細節不必深究，這只是個例子而已[[84]](#_84_52)），就會讓人得到寬慰。也許這些改善的前景足夠穩定和具體，因此我們可以設想有宗教信念是先天的。假設宗教信念最基本的形式類似于“上帝存在”或者“有超自然的力量作用”。這樣一來，按照坎貝爾的想法，有這種宗教信念就會是證成的。

以上的思考表明，我們迫切需要做出一個區分。我們要區分的是信念得到證成的兩種方式：在工具意義上證成和在認識上證成。對于某個人而言，如果一件事情能幫助她實現自己的目標，那這件事就是在工具意義上證成的。我們通常把這種證成的模式應用于行為，但將其應用于信念也沒有什么不一致的地方。如果凱特相信她是世界上最頂尖的網球選手，而且她覺得幸福（長期的、真實的幸福），那么對她來說（如果幸福是她的目標之一），這一信念在工具意義上就是證成的，哪怕她其實離頂尖水平還差得很遠。這個例子讓人覺得有些不對勁的地方在于，一般而言，我們提起信念的時候說的不是這種類型的證成。要是你聽到有人說，“凱特相信她是世界上最頂尖的網球選手，這個信念是得到了證成的。”你要是把這句話的意思理解成凱特的信念對她有利，那就很奇怪了。相反，你大概應該會覺得這話的意思類似于“她的信念是由一種對已有證據靈敏的過程形成的”，或者“她的信念是一種可靠地產生真信念的過程的產物”，或者更一般地講，“她的信念符合某些適當的認識標準”。這就是認識上的證成。假設適當的認識標準包括對證據的敏感性。那么，如果凱特沒有見識過證明她實力超群的證據（遺憾的是，事實上她見識過很多相反的證據），但依然堅持她的信念，而且這讓她感到更加幸福（長期的、真實的幸福），那么她的信念在認識上就沒有得到證成，雖然在工具意義上得到了證成。很多信念在這兩種意義上都沒有得到證成；在認識上沒有證成的信念，常常導致目標無法實現，因而在工具意義上也沒有得到證成。

坎貝爾將會如何回應這些批評還是比較清楚的。他會抱怨說，我設想的先天宗教信念的例子，和先天道德的情況有一個重要的分別。他會說，區別在于道德不屬于用來“指稱或對應于道德事實的領域”（21）的信念。道德信念的特殊之處不在于其主題，而在于“它們本身會引起某些行為或不作為，在缺乏這些動機的時候會導致內疚，對于富有這種動機的他人會產生敬仰和尊重，還會導致如下想法：這些動機足夠重要，從而有必要對缺乏它們的人實施懲罰。從這個角度看，道德信念本質上是按照某些規范去思考，去感受，以及去行動的。”（同上）他在文章第一段，就加入了這種非認知主義的理解，這樣一來，坎貝爾實際上使認識證成變得不適用于道德，也讓工具意義上的證成（OMA所提供的那種證成）變得恰當而且自然。我想提出異議的就是這個假設。坎貝爾說，他接受了非認知主義的理解，是因為不想因為偏向認知主義而回避問題，但是因為他接下來提出的論證的合理性完全依賴于對道德的非認知主義解釋，事實上他在相反的方向上同樣回避了重要的問題。[[85]](#_85_52)

如果有人像我一樣覺得純粹的道德非認知主義錯得無可救藥，那他就會覺得坎貝爾的論證不能令人信服。我們可以承認，也許他確實找到了一種證成道德的方式（他也許證明了具有道德“信念”是種有益的策略），但我們應該反駁說這種意義的證成沒多大意思而且可能是誤導性的。如果我們認為道德信念確實是信念，不僅是“去思考，去感受，以及去行動的傾向”，那么我們就會覺得道德信念同前面說的宗教信念同屬一類。說相信上帝能提高群體里每個人的幸福因而是證成的，和說基本的道德信念因為同樣的理由也是證成的，都一樣誤導人。畢竟，無神論者雖然認為宗教信念是錯誤的和/或沒有證成的，也常常愿意承認這些信念曾經或者現在是對社會有用的。坎貝爾的說法，完全沒有辦法阻止道德懷疑論者對于道德持類似的態度，視其為有用的虛構。坎貝爾就好像是在說，“有用？這意思就是它得到了證成。”而道德懷疑論者則會答道，“就算是吧，但它仍然是個有用的虛構。”

## 5.6　從演化的角度維護道德：丹尼特

丹尼爾·丹尼特在他的長篇杰作《達爾文的危險觀念》（Darwin’s Dangerous Idea,1995）的結尾部分討論了道德自然化的計劃前景。或者說，至少從章節標題來判斷，他討論的是這個問題。但是我不得不承認，在多次閱讀“倫理學可以被自然化嗎？”這一節之后，我還是不確定丹尼特的答案究竟是“可以”還是“不可以”（還是“這個問題本身就有問題”）。他先討論了動搖了傳統規范倫理理論的一些實踐問題。功利主義者遭遇的挑戰是把所有的后果都納入計算；康德主義者的難題是找出要進行普遍化的準則。我們在此無須斤斤計較（但是我很懷疑功利主義者和康德主義者在這些常見問題面前就此繳械投降）；他總體上的觀點是，生活非常復雜，而那些著名的倫理理論似乎太過簡單。這一節就這樣結束了。至于這個總體上的觀點如何有助于解答原本的問題，我還是找不到頭緒。

丹尼特接下來討論了我們的決策如何被簡單現成的啟發式過程所控制。對于解決時間緊迫的決策問題而言，這些過程相當不錯，雖然不能確保每次都得出最佳答案。他在此表達的觀點（也是書中的主要觀點之一）是，這些滿意型[[86]](#_86_52)過程貫徹始終：“……甚至不僅限于決策者固定的生物設計結構，還延伸到自然母親在設計我們和其他生物的時候采納的設計‘決策’”（1995：503；下文標出的頁碼都是此書的頁碼）。丹尼特把這一合理的現象稱為“滿意型的根基地位”。接著就是論證里的重要環節，即他從“是”滑向“應該”的過程。他進行得如此迅速和流暢，乃至很容易被人忽略，而且，這個過程幾乎去完全是借助羅杰·維特海默1974年的論文實現的：

因此，我們思考什么和如何思考，都是理性原則（即我們應該思考什么和如何思考）的證據。這種關系本身就是理性的方法論原則；讓我們稱之為“事實規范原則”。每當我們嘗試決定應該思考什么和如何思考時，我們都（暗地里）接受了事實規范原則。因為在進行嘗試的過程中我們就必須思考。而且在嘗試的過程中，除非我們認為思考的內容和方式是正確的（因此構成了我們應該思考什么和如何思考的證據），否則我們最終就無法決定我們應該思考什么和如何思考。（504）

對于維特海默的有趣主張，可以探討的有很多，但是這里不是合適的場合。關鍵在于，丹尼特借此從前面的描述性討論得出了一個評價性的結論：大致而言，因為我們自身的限制，這一被啟發過程控制的決策是我們所能做的最好決策。他隨后的工作是深入細節，提出能使這種決策得到調整修正和總的來說得到提高的一些具體方法。對他的論證而言，重要的是“對話終結者”的必要性。“對話終結者”這些考慮因素，出現在個人和人際間的決策過程中，是為了使這一過程達到有效的終點，從而避免該決策機制不停地尋求進一步的理據，無休無止地琢磨進一步的考慮因素。在丹尼特看來，道德價值、原則和命令就是最典型的對話終結者。他的觀點有個很有意思的地方，即為了讓我們的判斷要能夠有效地扮演這種角色，我們在使用道德判斷時一定不能將其視作對話終結者；相反，我們必須視其為表達了真正的、毫無疑義的、無須贅言的道德考慮。丹尼特承認，他的立場表明，“伯納德·威廉姆斯所說的社會“透明”理想——‘社會的道德機構能夠起作用，不應該依賴于社群成員對其如何作用的誤解。’[[87]](#_87_51)——是一個在政治上我們可能無法達到的理想”（509）。要能有效地起作用，道德思維的真正本性必須是晦暗不明的；我們在進行慎思的時候必須隔離對道德真相的理解。這種關于道德思考的用處（關于道德如何終止深思熟慮）的觀點，跟我在4.2節的論證相當吻合；事實上，到目前為止簡述的丹尼特的論證，幾乎沒什么好反駁的。我們現在必須要提出的問題是：“這個論證成功地維護了道德嗎？”

丹尼特明顯認為答案是肯定的。我之前承認了，我不知道他是否認為有哪種類型的道德自然主義被證實了。早先他說道，“倫理學必須以某種方式以對人性的理解為基礎——基于對人是什么或者可以是什么，以及對人想要什么或者想成為什么樣的了解。如果這就是自然主義，那自然主義就不是什么謬誤。”（468）說得是很對，但這不是自然主義！（對方一樣可以說：“如果道德自然主義只是說有的東西是粉色的，還有黃花邊——那為什么要為道德主義歡呼呢？”）道德自然主義是關于道德價值是否存在和道德判斷是否為真的一種主張（參見5.1節），而對于這些問題，就我所理解的，丹尼特是沒有表態的。

對于丹尼特成功維護了道德的提議，我將提出兩種批評。

我的第一個批評可以簡明扼要，因為有了上文對坎貝爾的討論，這種批評應該并不突兀。如果我對丹尼特的立場理解得沒錯，那他向我們證明了做出道德判斷這種活動扮演著極為重要的實踐角色。或許他還揭示了，為了從道德判斷中獲益，對于道德判斷如何起作用，判斷者本人必須不能想得太仔細（她必須有點“健忘”，且不會過度反思）。顯然，這無非只是工具意義上的證成，而且這同道德判斷實為錯誤，以及/或者道德判斷在認識上沒有證成，都沒有任何矛盾。相信某些行為會激怒紫蜥蜴大神，相信接觸某些物品會污染你的靈魂，或者相信有的女人與魔鬼同謀，也許這些信念作為對話終結者都很有效果，但與此同時作為對世界的描述也都是大錯特錯。至于如何區別道德判斷和以上這些明顯錯誤的想法，丹尼特未置一詞。即便為了獲得相應的利益，使用道德判斷的行為者不能對她的作為想的太仔細（即不能把相關的價值視為工具價值），即便那種價值依然是工具的價值。而且，行動者眼中的某個道德上被禁止的行為，即便出于實質或理性的原因確實不好，這并沒有證明有關該行為的道德判斷是真的或是在認識上證成的。出于同樣的理由，在某人眼中會激怒紫蜥蜴大神的行為，也許都是不審慎務實的行為，也就是說，就所有這些行為而言，“你不應當那樣做”（在此是務實意義上的“應當”）都是真的，但是這并沒有證明當“應當”的意思是“紫蜥蜴大神所命令的”時，“你不應當那樣做”這句話是真的或者是在認識上得到證成的。（有的批評者論證說丹尼特其實相信的是道德虛無主義，參見Rosenberg and Sommers 2003。）

我并不是說道德判斷的對話終結者身份蘊涵著它們都是假判斷的意思，我只是說，這一身份跟它們都是假的在邏輯上完全是融貫的。某些真話也可以作為對話終結者起作用。一般而言，“但那會害死我的！”這種考慮因素會終結對于某件事情進一步的慎思，但也許那件事確實會害死說話人。所以問題就成了，道德判斷是真的對話終結者還是假的對話終結者？如果是真的，那它們就不只是對話終結者；正如關于什么會害死人的信念，道德判斷也可以在仔細的、透明的慎思中起作用（即扮演非對話終結者的角色）。如果這是丹尼特想接受的觀點，那他的理論就有著極為明顯的問題，因為對于是什么樣的事實能使得道德判斷為真，或者道德判斷是如何在認識上得到證成的，或者它們在對話終結者之外怎么扮演其他角色的，他都未置一詞。（比較一下，對于“那會害死我的”這個命題，以上的任務可以如何完成）。其實從元倫理學的角度看，比起指出道德判斷也可以作為對話終結者起到作用，對這些問題的回答要有趣得多。因為對于這些問題的討論完全付之闕如，我們會覺得在丹尼特看來，道德判斷只是對話終結者而已，但那樣一來，道德判斷就完全沒有理由要求被嚴肅地當作是真的。按照這種解讀，丹尼特提倡的是一種道德虛構主義（moral f ictionalism）。這種立場由以下四個論點的組成：（i）道德判斷事實上不是真的；（ii）我們可以知道它們不是真的（比如通過哲學研究）；（iii）它們還是有用的；（iv）然而只有當我們在日常生活中將其視為真的，它們才會有用（更多有關道德虛構主義的文獻，參見Joyce 2001a，2005；Kalderon 2005；Nolan et al.2005）。我不是批評丹尼特提倡道德虛構主義（如果他確實擁護這一立場），因為我自己其實很喜歡道德虛構主義。我目前的主要反駁意見是，如果丹尼特的觀點算是對道德的維護，那也只是在工具意義上的維護。

我的第二個批評是，我很難理解丹尼特表達的、關于對話終結者的看法（也許可以算是道德虛構主義的觀點）和他的書最后大量貌似“直截了當”的道德說教之間的關系。最后一章充滿了道德沉思、建議和懇求。他讓我們反思為什么殺死一只牛要好過殺死一只禿鷲（513）；他告訴我們瀕臨滅亡的語言是“值得保留的、珍貴的人類創造”（514）；他告知我們奴隸制和虐待兒童是“不可接受的”（516）；他請我們跟那些公開反對針對薩爾曼·魯什迪的“不道德到無法形容的”追殺令[[88]](#_88_53)的阿拉伯和伊斯蘭作家“攜起手來”（517）。在最后的高調宣言中，他告訴我們，“這個世界是神圣的。”在丹尼特所揭示的不透明性的背景下，我們應該怎么理解這些說法呢？是否它們又只不過是一些對話終結者，來讓作者和讀者避免無休止的思考？還是說，我們現在已經被很快地擋到了晦暗不明的帷幕背后，從這個角度看，這些想法（其實只不過是對話終結者而已）不再被當成對話終結者，而成為真正的、有約束力的實踐考慮因素？這實在令人困惑，而我也因此不能不懷疑，丹尼特是在透明和不透明的問題上一再反復，在他的頭腦告訴他的（即“奴隸制是錯的”這句話的作用是終結對話）和他的內心告訴他的（“不對，奴隸制確實是錯的！”）之間輾轉不定。

我們想知道，按照丹尼特的觀點，我們究竟能不能發現看穿道德終結者的本來面目。在決策的過程當中，我們當然需要相信道德規定表里如一（即相信它們具有真正的、無疑的實踐影響力的要求），但當我們靜下心來的時候，是否必須也繼續這樣想呢？丹尼特面臨著兩難的處境，因為無論如何回答，都會導致問題。如果他的理論意味著人們永遠不能從晦暗的帷幕背后往外看，那這理論就是有問題的，因為根據假設，在這帷幕背后的人們壓根沒人會同意道德判斷的功能是終結對話：要是讀到某個哲學提議，后者指出“奴隸制是錯的”這個判斷的價值在于它能有效充當一個方便的對話終結者，所有讀者都會抗議說，“那不對，奴隸制確實是真的錯了！”進一步地，我們也會懷疑為什么就丹尼特自己能有幸站在道德框架外面，見識到它的本來面目。因為有這些問題，我們不得不認為丹尼特的意圖是想說，通過進行抽象的沉思（比如思考《達爾文的危險觀念》末尾的論證），我們能夠承認，道德判斷確實只是方便的工具。畢竟他似乎可以這樣宣布，而他當然也希望我們能夠贊同。兩難處境這一端的尷尬之處在于，看清了這種對話終結者的本來面目，也就使其喪失了發揮作用的能力。如果丹尼特想要讓人們保持他們的道德信念（他最后的說教似乎表明他似乎確實希望如此），那么也許他本應該把他有關道德對話終結者的觀點按下不表，或者不要給出那么令人信服的論證！

我的意思不是說這個兩難是致命的處境。道德虛構主義也許就是緩解這種困境的好辦法。然而，展開一套虛構主義的論據（即讓上述（i）至（iv）不再顯得自相矛盾），需要一定的細致區分和注意力，而丹尼特的論述并沒有體現出來；事實上，他似乎根本沒有意識到上述兩難處境的存在。他似乎疏忽了，對于“這個世界是神圣的”這句真誠的結語，書里不久之前的內容會使得讀者做出如下回應：“那么，這只是個對話終結者，對吧？”他也沒有注意到，如果答案是肯定的，顯而易見，接下來的問題就會是：“那么，這個世界究竟是不是神圣的？”

## 5.7　從演化的角度維護道德：凱斯比爾

最近，威廉·凱斯比爾（William Casebeer）在演化生物學的基礎上為道德自然主義進行了辯護。更具體地說，他辯護的是“實用新亞里士多德主義的德性論”（“pragmatic neo-Arisotelian virtue theory”）。根據新亞里士多德主義對于德性的解釋，道德評價的主要對象不是行為而是人的性格。這種解釋也構成了凱斯比爾的論證基礎。對亞里士多德而言，“一個好人”就類似于“一棵好橡樹”、“一個籃球好手”（也可以是“一個好殺手”），大意是指相關類型的典范[[89]](#_89_52)。如果相關的類型是功能性的，那么該類型的典范就是很好地實現這種功能的東西。亞里士多德論證道，正如基薩拉琴師的功能是彈奏基薩拉[[90]](#_90_52)，因而好的基薩拉琴師就是基薩拉彈得好的人，同樣，明白了人的功能（the ergon）是什么，我們也就應該能夠發現怎樣算是一個好人。為了確定人的功能，亞里士多德采用了他標志性的方法，即先明確相關的種（這里相應的種是動物），再找出在這個種內與眾不同的特征（即屬差；這里的屬差是理性）。亞里士多德認為，人的功能是“行為和靈魂的活動都遵循理性原則的生活方式”，而很好地實現這一功能，就等于是遵循德性而生活（“如果不只有一種德性，那就遵循最好的和最完整的德性”）[[91]](#_91_52)。這些德性既有智性上的，也有性格上的，而后者（勇氣、友誼、大度等等）被后亞里士多德主義的解釋者稱為“道德德性”。

亞里士多德的論證的主要問題一直都在于，它堅持掌控世界的原則是目的導向的。在《物理學》一書中，他為目的論的宇宙觀提出的論證既簡陋又令人困惑。首先，亞里士多德堅持了一個目的論和偶然性的奇怪二分法：任何有規律地和可以預料發生的事情，比如冬天的雨（但夏天的雨就不行），一定是為了某種目的才發生的。隨后就是令人瞠目結舌的宣言：“有理智的行為是如此，當然在自然界中也是如此；在自然界中是如此，類似的，任何未經干擾的活動也都是如此。”（我頭腦里的紅筆只能在這句話邊上寫下“真是這樣？？”以示震驚。）亞里士多德把與此對立的立場（即自然發生的事情沒有目的）歸于古代思想家恩培多克勒，而后者的觀點比當時任何人都更接近達爾文式自然選擇的理論雛形[[92]](#_92_52)。這多少有些反諷，因為許多當代的思想家，在正確地否決了古怪的亞里士多德式自然目的論之后，又想借助達爾文主義的生物學來給我們的科學世界觀添加這種目的論的部分元素（例如，參見Allen et al.1998）。自然選擇不是趨向一定目的的過程，但是由它產生的結構和機制，從自然主義的角度看來，用自然目的和功能來形容似乎是正當合適的。當然，這種新目的論不會把功能賦予冬雨，但會賦予眼睛、心臟、魚鰭、鞭毛、根莖、膽囊等一定的功能。帶來了一套對于功能屬性的解釋，按照這套解釋，功能屬性是“徹頭徹尾的自然的、不奇怪的”（2003：49）。他認為，有了這套解釋，就可以用好的達爾文式目的論替換不好的亞里士多德式目的論，從而拯救亞里士多德對于人類德性富有洞見的理論。而這樣一來，他認為，就給從演化的角度對道德的維護鋪平了道路。

讓我們姑且同意，演化生物學使得自然功能的說法以及與之伴隨的規范性語言，在科學上都是正當的。現代生物學指出，心臟的功能是泵血，而健康良好的心臟就是泵血泵得好的心臟。它甚至還會允許我們說心臟“應當”泵血，而這不僅僅是預測性的“應當”。但這種規范語言跟道德語言有多接近呢？它是從正確的地方找到規范性的嗎？而更重要的問題是，這種規范性的類型是合適的嗎？

關于它是否找對了地方的問題，我們先來看一下亞里士多德常舉的那些有功能的事物的例子。在《尼各馬可倫理學》一段重要的文字里，他給出了兩類例子：有社會角色的人（基薩拉琴師，雕刻家和木匠）以及身體的部分（眼睛、手、足）。這兩種事物都處于賦予功能的環境中：基薩拉琴師在社會里有一定的職業，眼睛則是在身體上有一定的分工。亞里士多德試圖把這些例子拓展開來（給一般意義上的人賦予功能），可這沒有說服力，因為此時就沒有了賦予功能的環境。所以亞里士多德不得不依賴于他那（在現代人看來）古怪的目的論物理學，來確保他的論證能成功。一旦我們否定了這種物理學，那么即使我們不排斥說木匠和眼睛具有功能，我們還是可以拒絕承認人類（就其為人類而言）具有一定的功能。現在的問題是，凱斯比爾那種在科學上值得尊重的目的論是否在這一點上做得更好。雖然演化生物學允許我們有意義地說起眼睛、手和足的功能，但它并沒有明顯地給人類也賦予功能。因此，那些因為生物學而變得正當的規范性話語，比如某個X的功能是ϕ，某個好X做ϕ做得不錯，X應當ϕ，要是X是“心臟”或“眼睛”，都是可以的，但我們沒有理由將其替換成“人類”。怎樣從“喬的心臟應當泵血”得出“喬應當遵守諾言”，都是有問題的。

可能有人會反駁說，我把生物學里的規范性理解得過于狹隘。也許我們無需從“X的功能是ϕ”直接推論出“X應當ϕ”。也許我們從生物學中得到的，本應是關于生物體怎樣才能狀況良好，怎樣才能繁榮發展（f lourishing）的某種值得尊重的、客觀的觀念。也許有人會認為，這就等于是科學發現了生物體的目的是什么。一旦找到了這種目的，就可以得出以假言命令形式出現的規范性話語。如果人類的繁榮要求（例如）在社群中和諧生活，那么只要違背承諾就會損害人的社會身份，那么我們就可以堅持“你不應當違背承諾”這條假言命令是適用的。對于聲稱社群對自己一文不值的罪犯，我們可以說“同社群和諧生活是你的目的之一（價值之一），不論你是否明白或是否喜歡。”因為這些自然目的在某種意義上是無可避免的，相應的假言命令也是無可避免的；這所導致的規范性，也許有人會覺得，就會有我們希望道德所具有的實踐力量。

會嗎？這里有必要重溫一下本書2.4節討論的內容，即假言命令無法體現道德實踐性。我們首先來考慮一下道德思維的一個核心概念：違規（trasgression）。未能遵循假言命令不能被視作違規（否則就極為別扭）。要是某人因為冷而覺得不適，而關窗能讓她暖和，那我們可以說她應當關上窗；但是，如果她沒這么做，那么盡管她可能是愚蠢或不明智的，很難說她違反了規范。只要我們覺得道德系統應該能容納“犯罪的人應為其罪行而受懲罰”，假言命令組成的系統就無法滿足這種要求。這里根本的問題在于，沒有實現自身目的的人主要傷害的是她自己，但都圍繞著自我傷害而構建的價值系統不大像是道德系統。在這樣的系統中，懲罰就成了因為某人傷害了自己而對她造成進一步傷害的怪異制度。按照這種觀點，開膛手杰克應當引起我們最深切的同情，因為他殺害那些女人的真正錯誤之處，在于他因此嚴重破壞了自己的繁榮興盛。（按照這種觀點，抓到他之后我們應該嘆息道：“可憐的家伙。”）[[93]](#_93_52)

演化生物學也許會準許我們說某顆心臟是好的，因為它泵血泵得好，但這么說并不是真的在贊美這顆心臟，就好像我們承認某個殺手職業手段高超并不是在贊美他冷酷高效一樣。因此，按照這種邏輯，如果演化生物學也準許我們說某個人是“好的”（意思是她是人類繁榮的典型代表），這肯定也是在“非贊美”的意義上使用這個規范詞語，而不是道德意義上的用法。演化生物學也許會準許我們說心臟應當泵血；但它不會容許我們說心臟是被要求泵血的，說它有理由泵血（這不能跟說它泵血是有緣由的混為一談），說有一定的實踐命令要求心臟（不論它愿意與否）繼續泵血，說心臟要是沒有泵血就違規了，所以應得懲罰，或者說不存在什么讓心臟停止泵血的壓倒一切的理由。

讓我們停下來再考慮考慮以上最后一句話。錘子的功能是釘釘子，但用錘子當門擋也毫無問題。撐著打開的門時，錘子還是保留著本來的功能，但該功能現在沒有實現一點也沒有關系。演化的功能憑什么就與此不同呢？我聽說，人類巨大的臀大肌演化出來，是為了提高我們祖先扔東西的能力（Bingham 2000）。先放下你的懷疑，假設這是真的，并且假裝直到不久之前，這確實是我們的祖先用這塊肌肉做的事。但設想一下你恰好不喜歡扔東西（也許你小時候被一個棒球投手造成了精神創傷），而且你決定再也不扔任何東西。你做錯了什么嗎？你違犯了自然制定的規范了嗎？也許有人會做肯定回答，因為他預設了不使用自然功能很可能損害健康或者導致不快樂，但這樣一來，其實是不健康或不快樂使得這種決定錯誤，而不是沒能實現某個功能本身。再考慮另外一個例子：假設演化生物學確實以某種方式支持了“杰克不應當殺人”這個規定。這個“應當”的要求為什么能壓倒一切可能出現的、與之對立的“應當”要求呢？假設杰克可以從殺人得到巨大的愉悅，而獲得愉悅也是杰克選擇的目的之一；于是就有一個對立的假言命令：（為了感到愉悅）“杰克應當殺人。”有什么根據說前者壓倒后者呢？有什么根據說，對杰克而言，前者具有而后者缺乏某種“要做到”的性質呢？也許有人會回應說，前一個命令的基礎是杰克的生物機理，或前者不像后者那么偶然，或者甚至說（相當極端的說法！）前者是從杰克的人類身份的某些本質得出的。然而，前一個命令的各個特征（或其組合），都不足以使其擁有所需的勝過一切其他命令的性質。持相反觀點的人還是可以說，最具壓倒力量的命令來自一個人最衷心在意的目的（就杰克而言，那是他惡劣的欲望），而且這么說表面上的根據也一樣有力。要是我們從古人類的遺跡中發現，造成道德感的選擇壓力，主要來自凝聚族群以便跟其他族群爭斗的需要（這可能性不大，但姑且這么假設），顯然，沒什么人會因此斷定我們現在要特別青睞和實現這一功能。[[94]](#_94_52)

在凱斯比爾眼里，他的主要論敵是約翰·麥基。麥基有一個著名的論點，即道德話語跟占星術或有關女巫的言論一樣，對世界的描述是錯誤的。思考麥基的觀點對我們很有益處，因為我之前對丹尼特和凱斯比爾的反駁，可以簡單總結為：即便他們的論證是可靠的，麥基對于道德的論斷依然可能是正確的，因此這就足以表明在他們手中道德還是“未得維護”。麥基的一個中心論證指出，道德屬性要是存在的話，在形而上學上會是非常古怪的，“……它們會是非常奇怪的一種物體，或性質，或關系，完全不同于宇宙里的其他任何事物。”（1977：38）為了說明它們會是多么古怪，麥基用了柏拉圖對至善形式的解釋作為例子。據說光是認識到某個東西有了這種形式（即是善的），就可以自動促發對這個東西的追求。在柏拉圖看來，這種善本身就包含了某種磁力。凱斯比爾自己設定的目標，就是提出一種絲毫不“古怪”也絲毫不像“幽靈般”的道德屬性，以此駁斥麥基。

但是，柏拉圖的形式理論不是麥基唯一的例子。他還提到了薩繆爾·克拉克，后者在18世紀早期就論證了（用麥基的話說），存在著某種“情境和行為之間的必然的適當關系，因此一個情景本身就包含了對于某一行為的要求”（40）。需要注意的是，這里并沒提到說人的動機是被道德善促發的（compelled）；只是說動機是被要求的。這表明，麥基所謂的“古怪”有另一種起源，而且有人論證說這更好地解釋了他的目標（Garner 1990）。古怪的不是內在的動機影響作用，而是其內在的行為導向性。麥基在另一本書中也描述了這種古怪：“……說［客觀的規定］本身內在地引導行為，意思就是說這些規定給出的做什么或不做什么的理由，無關于行為者本人的欲望或目的”（1982：115）。這更好地刻畫了麥基眼里的真正的古怪之處：無論某人的目的如何都能為其提供行為理由的性質。麥基認為，世界上沒有東西具有這種性質，而這就導致了他的道德懷疑論。

至于這些提供理由的性質，是否真如麥基所說的那樣是古怪的（是否也因此如他所說是不存在的），不是我現在關心的問題（我在第六章會探討）。這里討論他的觀點是為了表明，一旦放下他那個誤導人的柏拉圖式形式的例子，更多關注他關于要求和理由的說法，我們就可以看到他對于道德本質的見解一點也算不上出格。凱斯比爾寫道，“麥基的意思是說，價值必須包含著它自身的動機結構，而這看起來是很奇怪的，因為‘動機’并不是在世上四處漂浮，等待著被道德行為者感知到”（41）。如果這確實是“麥基的意思”，那看起來是很奇怪，而凱斯比爾覺得無論他自己還是別人都沒有義務認同這一點，他的想法也會是正確的。然而，這真是麥基的意思嗎？假如我們把凱斯比爾引用的文字里的“動機”換成“要求”（“demand”）：“麥基的意思是說，價值必須包含著其自身提出要求的結構，而這看起來是很奇怪的，因為‘動機’并不是在世界里四處漂浮，等待著被道德行為者感知到。”我認為，這樣一來“麥基的意思”立刻顯得毫不奇怪了：日常道德話語似乎是堅信“在世上四處漂浮”的要求（如果這指的是那些其權威性不來自我們的要求；參見2.4、4.5、4.5節。）確實存在。如果這是古怪的，那這也是所有的道德自然主義者都要接納的古怪，否則他們所提議的不古怪屬性就跟道德沒有什么相似之處了。但是，我們之前也看到了，這跟從演化生物學中得出的規范性相差太遠。正如心臟的功能是泵血這一事實不意味著存在一種讓心臟泵血的要求，出于同樣的邏輯，賦予人類一定的功能也不意味著存在著任何實踐要求。正如某殺手的功能是殺人這一事實不意味著他有任何理由那樣做（至少沒有對于他的慎思而言有一定分量的理由），出于同樣的邏輯，賦予人類以一定的功能，也不意味著有任何實現該功能的實踐理由。

在書末（第153—155頁），凱斯比爾探討了他的理論是否只給我們提供了“孱弱的規范性”。然而，在他對該批評的回應中，他把對方的觀點理解成了堅持道德規范必須是絕對的和可以確定知道的——但這些都不是我所討論的道德規范需要滿足的條件。我們姑且同意道德規范可能不是絕對的，并且我們對其的認識是可能出錯的。盡管如此，要是有一個價值系統，其規范：（i）不產生要求和實踐理由；（ii）很容易被對立的考慮因素壓倒；（iii）沒能遵循這些規范也算不上是違規或不應受懲罰的；（iv）宣布某人遵守了這些規范也不算是贊揚，那么，這個系統就太過于孱弱了，因此也就不大可能被當成是道德。

## 5.8　結論

以上四種用演化假說來維護道德的嘗試各有不同，但我們也見到了一些反復出現的主題，從中我們可以得出更具一般性的結論。坎貝爾和丹尼特試圖從工具的角度維護道德，但就算他們成功了，道德在更重要的意義上還是“沒有得到維護”。哪怕他們說的都是對的，道德依然可能只是某種有益的錯誤：在實踐上有用處但全都是錯的（這是無神論證對宗教話語常有的態度，似乎也是麥基對道德話語的態度）。由于假設了道德判斷完全是非認知性的東西（我們在第二章已經看到了該假設的某些問題），有的人覺得工具性的證成是維護道德的合適方式。第二種錯誤（體現在理查茲和凱斯比爾身上）是錯認了演化論里的“應該”的類型。這種錯誤之所以產生，是由于忽視了“一個規范框架要在我們的生活中起到我們要求道德所扮演的角色，就必須具有那些特點。”的問題。在本章的結尾，讓我們對這兩個問題再各做一點討論。

首先，也許有人覺得，對于道德感的演化最合理的假說，事實上提高了純粹非認知主義的可信度。例如，基徹爾（1998，2005）就論證道，如果道德的譜系學意味著道德判斷的功能是協調人類的社會行為，而且如果這一功能的實現是通過“擴大我們的心理利他主義的傾向”（2005：178），那么非認知主義的前途看起來一片光明。他的論證似乎是，如果道德判斷的功能是增強我們情感生活的某些方面，那么“道德判斷的表明形式欺騙了我們，（因為）我們其實不是在說出簡單的陳述句，而是在表達情感反應”（同上：175）。但這里的要旨在于，我們不能混淆了道德判斷的演化功能和語言功能（基徹爾似乎混淆了二者）。假設道德判斷的演化功能確實是產生或增強某些親社會情感。我們必須要問：“它是怎么實現這些功能的？”一種答案或許就是：道德判斷促使說話者（及其聽眾）把世界看成是包含有某些獨立于權威的實踐要求，而這樣看待世界可以對他們的情感產生想要的效果。如果事情果真如此，那么當說話人說“你不能那樣做”，她就確實是在對世界做出斷言，而不僅僅是在表達感情。也許她所斷言的是錯的，但這無損于其斷言的地位。（這一點類似于我在4.4節就認知主義和非認知主義提出的觀點；同樣可以參見Joyce 2001a：14—15。）我們可以再類比一下有關先天宗教信念的假說。（為了簡單起見）假設上帝存在的信念是先天的，而且只是由于該信念可以幫我們的祖先減輕焦慮（使其更加高興）。但由此很難推論出，說出“上帝存在”這個句子不是純粹的斷言，或者說出這個句子就只是在表達說話人的高興而已。

規范的演化倫理學家所犯的第二個錯誤，是誤認了演化過程中的規范性類型。理查茲和凱斯比爾（以及我在別處批評過的其他作者[[95]](#_95_51)）認識到需要的是認識的而非工具性的證成，但沒有意識到道德價值和判斷所充滿的特殊的實踐力量（為了達到我們要求的具體目的，它們必須具有這些力量）。由于缺乏這種意識，他們也錯認了（據說是）我們用來獲得和提高道德信念的認識手段。因為這些理論家們關注的規范性類型不對，他們對于我們如何發現道德真理所持的見解頗為古怪，這也沒有什么出奇的了。在本章的最后，讓我來提請讀者注意一下演化的道德自然主義在認識論上的怪異之處。這并不是作為毀滅性的論證而提出的；但是，我認為這個簡單的論證揭示了這種理論的一個明顯違反直覺的后果。

如果道德判斷是從有關演化的事實得到的認識證成，那么有關演化的，意料之外的新發現就可以（有些時候還應當）改變我們的道德看法。例如，假設按照迄今為止由考古證據得出的最好理論，我們認為人類演化過程中相關階段的社會條件是這樣或那樣的，這讓我們接受了有關人類進化的X理論。根據演化的道德自然主義，這應當讓我們暫時支持某些具體或一般的道德主張，比如“在妊娠的后三個月墮胎，有時候是允許的”，“撤銷第三世界的債務在道德上是可欲的但不是義務性的”，“應該把埃爾金大理石雕歸還給希臘”。但是再設想一下，后來出現了一些意料之外的古生物學證據和來自史前遺址的證據，對于我們祖先的社會組織，這些證據意味著一些非常不同的結論，于是我們修改了對于人類演化相關方面的觀點，采納了理論Y。根據演化的道德自然主義，這可以要求我們修改當代道德觀點的部分乃至全部內容：我們過去以為撤銷第三世界的債務是可欲的但不是義務性的，在原則上，今天我們可能會覺得那是不可欲的但也不是被禁止的，并相應改變我們的政策。如果一種理論允許當今世界復雜的道德爭論可以（哪怕只是在原則上）通過挖掘非洲的土地得到解決，或者是主張由于我讀了《自然》雜志上有關新發現的人科化石的文章，那因為違背了上周對我岳母許下的諾言而產生的負罪感就會瞬間煙消云散，那么這種理論當然是極難信以為真的。盡管我對考古學家非常尊重，我不認為我們評估道德真理的時候應該求助于他們。

# 第六章　用演化拆穿道德

## 6.1　譜系揭穿真相

任何信念都有一定的來由，所以只有荒謬絕倫的理論才會指出，光憑對于信念起源的了解，就能自動地削弱對于該信念本應持有的信心。但是，顯然在某些情況下，對于某個信念起源的知識是可以削弱原有信心的。讓我們來做一個有點愚蠢的思想實驗：設想世上有種信念藥片，吃下去之后就會不可避免地產生某一個特定的信念（同時讓人忘記曾吃過它，并且為了安全起見，也會讓人全部忘記這種藥片的存在）。假設有一個藥片能讓你相信拿破侖打贏了滑鐵盧之戰，另一個藥片則能讓你相信他輸掉了滑鐵盧之戰。再假設有兩種藥片都有同一種解藥。現在，想象一下，你一直以來都愉快地相信拿破侖輸掉了滑鐵盧之戰（就跟事實一樣），然后你發現了，曾經某刻有人偷偷給你下了“拿破侖輸掉了滑鐵盧之戰”那個信念藥片。現在的情況不是你了解到這種藥片的存在卻無法知道是否曾經吃過；我們設想的情況是，你不知怎么發現了你的信念確鑿無疑是該藥片的產物。這應該打破你對“拿破侖輸掉了滑鐵盧之戰”的信心嗎？當然應該。這一發現并沒有證明你的信念是錯誤的（雖然這里虛構的情景不同于現實世界，但也可能有拿破侖，滑鐵盧之戰，以及他輸掉該戰役的事件），但這一知識足以把你的信念放進可疑名單。關于一個信念的譜系學知識，只有在該信念蘊涵著相反的譜系歷史時，才能證明該信念是錯誤的。例如“沒有信念是先天的”這一信念，如果我們發現了它本身就是先天的，就可以證明它是錯誤的。但你肯定也注意到了，這樣的例子非常罕見[[96]](#_96_51)。我們的大部分信念都不大可能蘊涵著其自身由來的真相，相應地，譜系學的發現也不大可能直接揭示我們的許多信念都為虛假。但是譜系學的發現仍然有可能揭示了我們有許多信念其實都未得證成，而這種可能性幾乎同樣令人不安。在我們設想的例子里，知道了你的信念是一個藥片的產物，這就使得那個信念未得證成（按照有的認識論觀點，這也使得那個信念從來就沒有得到證成），而這也就相應地要求你不應當再持有該信念（也就是說你應該服下解藥），除非你能找到具體的證據來支持或反對該信念[[97]](#_97_51)。

人們可以不無道理地反駁說，就其對人類心理的可能影響而言，自然選擇的作用方式肯定跟這些神奇的信念藥片完全不同。

首先，反駁者可以指出，人類的任何性狀都不是“內置”的，也就是說，人類的任何性狀都不是無視環境因素而形成的；同其他所有的表型適應一樣，一個先天信念需要環境的影響才能顯現出來。但是，這一事實并沒有損害先天信念這個概念本身。需要環境條件的信念并不因此就等同于學到的信念，也不因此就成了由某種對證據敏感的方式產生的信念。做個簡單的類比，設想有某人因催眠術而產生了一個信念。催眠的過程常常包括有條件的指令，比如“當你聽到‘banana’這個詞時，你就會相信你是一只雞。”（真正的催眠可能不是這樣進行的，但沒有關系，這個歪曲的好萊塢式催眠足以表達要點。）在這個例子里，環境必須要提供一定的誘因，但這個誘因對于此人而言并未構成她是只雞的證據，而且她也沒有從“banana”這話里學到任何跟雞有關的事情。

其次，需要注意，本書沒有任何地方宣稱道德信念是先天的。即便有時候我們可以合理地猜想自然選擇對于信念的內容（也許是有關互惠的道德判斷）有直接的興趣，但是因為道德信念往往涉及的是特定的事情。（例如，我相信因為瑪麗在機場接了我，我應該回報她），要是還堅稱它們是先天的就完全不合理（因為對于瑪麗和機場，自然選擇知道些什么呢？）我在前文論證的假說沒有否認，文化學習對于決定一個人最終做出的道德判斷的內容起到了關鍵作用；我的主張只是說，存在著一個專門的先天機制（或者一系列的機制），其功能就是使這種習得成為可能。我的假說是，這種機制天生就準備好了用道德的規范詞匯來把世界分類；即便道德信念不是先天的，道德概念仍可能是先天的。

然而，其實以上觀察對于前文的懷疑結論沒有什么影響；我們可以改變類比例子的細節來與其保持一致。假設這些想象中的藥片不產生具體的命題信念，而讓你傾向于形成包含某一特定概念的信念（在正常情況下，這個概念不會出現在你的信念中）。因此，這個藥片沒有讓你相信拿破侖輸掉了滑鐵盧之戰，它只不過是個讓你形成以拿破侖為對象的各種信念的“拿破侖藥片”。沒有這個藥片，你本不會形成任何關于拿破侖的信念。我們無須太擔心還有什么因素決定這些拿破侖信念的具體內容；也許這些信念是隨機決定的，也許是某些環境誘因產生了這樣那樣的拿破侖信念。假設你又發現你確鑿無疑曾在多年前被人偷偷下了這種藥。這一發現動搖了你所有有關拿破侖的信念了嗎？當然是的。信念中的一個概念被動搖了，這個信念本身也就被動搖了。同樣，這沒有證明信念是錯的，但在你找到能證實或證偽你的拿破侖信念的可靠證據之前，你應該服下解藥。隨著解藥生效，你失去了全部有關拿破侖的肯定信念，于是你就可以用你未經污染的頭腦去探索世界，去發現你之前的信念有沒有確實為真的[[98]](#_98_51)。

這個虛構的情景是為了方便我們理解關于演化出的道德感的一個類似的認識論結論。把虛構里的拿破侖信念替換成道德信念，把信念藥片替換成自然選擇。這個論證指出，如果不是因為祖先的社會生活影響了我們生物機理，我們就不會有諸如義務、德性、財產、應得和公平這些概念。如果類比是合理的，那么一旦認識到了道德的譜系，我們就（在認識上）應該做類似于服用解藥一樣的事情：培養與這些概念的有關的所有正面信念的不可知論，直到我們找到了支持或反對這些信念的牢靠證據。要注意這個結論有多極端。這不僅僅是讓我們對墮胎的錯誤或取消第三世界債務的正確保持開放的態度；這是讓我們對無論任何事情在道德上是對是錯都保持開放的態度，是讓我們承認，用道德詞匯來描述世界，跟嚴肅對待占星或相信先人的靈魂仍在世上游蕩，可能都是大同小異的（約翰·麥基論證說事實確實如此）。

但是，這個類比確實是恰當的嗎？也許有人會反駁說，信念藥片的故事是刻意設定好的，故事里通過吃藥形成信念的過程，和該信念為真所必須的事態在世界上是否存在，兩者完全沒有關系。但也許道德的演化譜系并非如此；也許自然選擇的過程很可能產生真信念。本章的大部分內容都旨在證明這個反駁是不成功的。對于道德感的而言，我們沒有理由認為自然選擇所產生的信念很可能是真的。所以這個類比，以及相應的令人不安的懷疑主義結論，都是成立的[[99]](#_99_51)。

首先，有一件重要的事情值得注意：在這個問題上，我們對道德的演化假說可能是不同尋常的。對比一下另一種演化假說似乎也能合理解釋的心理學現象。有一些證據表明，自然選擇賦予人類了一個內在的簡單算術能力。（參見Butterworth 1999）為了簡明起見，我們可以解釋說，這就意味著我們對1+1=2的信念是先天的。說1+1=2是永恒和必然的真理，大概肯定不會有什么問題，因此自然選擇把這個信念內置于我們的大腦里也毫無風險（環境因素不會突然發生變化，導致1+1=3[[100]](#_100_51)）。那么，我們對于簡單的數學信念有了譜系上的解釋，這證明了我們擁有這個信念是未得證成的嗎？當然不是，因為要是不訴諸該信念之為真，我們就無法理解自然是怎么選擇這個信念的，也無法理解這個信念是如何提高繁殖適應度的。錯誤的數學信念不會有用。設想一下，你被三只獅子追逐，然后看到有兩只放棄了，于是你得出結論說現在慢下來是安全的。“1+1=2”本身為真，在任何解釋相應信念何以成為先天信念的合理假說中，都是基本的前提。以下是另一個對比，這次涉及的不是先天信念而是先天概念：假設對于人類而言，兒童的概念是先天的。要是這個假定能得到充分證據，就會讓我們關于兒童的所有信念未得證成嗎？答案依然是否定的，因為相關的譜系解釋肯定會涉及兒童的存在。只有在兒童存在的環境里，發展出一套預備形成有關兒童的信念的機制才是有用的；另外，將預備機制轉換成關于兒童的現實信念，這個過程所必須的環境“誘因”大概也得包括跟兒童的實際接觸。

現在，讓我們把這種思路運用到道德信念上。如果不訴諸對和錯的存在，我們還有可能解釋為什么形成有關對錯的信念對我們的祖先是有用的嗎？我認為，答案是一聲響亮的“極有可能”。本書的前幾章對于為什么形成道德對錯的信念對我們的祖先有系統性的益處，提出了一套復雜的闡釋。但那套闡釋完全沒有預設在祖先環境里存在著真實的道德對錯。不論我們是否假設對和錯的概念成功地指稱了世界里的屬性，也不論我們是否認為這些概念指稱失敗因而同女巫、幽靈這兩個概念的地位相同，我們的道德判斷演化假說都不會受到影響。數學的情況就不一樣了。如果有人蠢到開始懷疑1+1=2的真實性，那么對他而言，對于相信1+1=2何以能提高人類祖先的繁殖適應度的演化解釋，就不再顯得合理了。

這也同樣適用于進行科學研究的能力，后者在某種意義上也是生物自然選擇的產物，而且我們也運用這種能力來構想和檢驗我們的演化假說。這就避開了彼得·雷爾頓（Peter Railton 2000）對于用演化理論是否可能破壞倫理學提出的顧慮。雷爾頓宣稱，任何從“人類道德是演化的產物”這一經驗前提，得出結論“道德在某種意義上被拆穿了”的論證，都“把自己也砸入了之前敲進道德的那塊地”（2000：57），因為我們用以建立前提的能力本身就是演化的產物。但是我想提醒注意的是，（就像在算術的例子里）如果進行“科學研究”有關的人類先天能力之所以被自然選擇，跟它能夠產生某些至少同真理有正面關聯的判斷毫無關系，那我們就無法理解這種能力為什么會被選擇。因此，“從演化的角度拆穿道德”并不會以同樣的方式拆穿自身。菲利普·基徹爾也表達過類似的觀點：

……社會訓練……對于科學判斷（比如技術人員判斷有一個電子穿過了云室，生物學家看到有的細菌吸納了額外的DNA）很重要。難道我們不能說因為有了這種訓練，所以無論電子或轉基因的細菌在場與否，技術人員和生物學家都會做出他們事實上做出的判斷嗎？他們在這些情況下做出了這些判斷，是因為他們經過了訓練，會對特定的視覺刺激做出特定的反應。但是，就科學而言，即便非社會性的事實也對直接的判斷沒有作用，這些事實依然跟一個更深層的解釋性問題有關。為什么相關的觀察者要接受這些類型的訓練？要回答這個問題，我們必須回頭看到這一社會實踐的歷史，而這歷史的一部分內容就是某些程序因其能夠可靠地察覺到世界上的事實而被采納。但是如果我們對于道德實踐（我們在其中被社會化）的歷史提出類似的問題，我設想的譜系學就沒法給出類似的寬慰。在最開始的階段，道德的原型，作為簡單原始的規則體系，是為了超越早期人類艱難的社會聯系而引入的：這跟認知道德真理沒有關系。在接下來的階段中，也沒有必要假設道德真理在限定被采納的規范系統時起到了任何作用。成功的標準并不是準確地反映了世界，而是以能促進該規范系統本身傳承的方式提高社會凝聚力。（2005：176）

觀察到道德的演化譜系不同于其他情況（比如算術上和科學上的信念），即前者無須假設信念為真，這是很重要的。但這還不足以證明，對于前者，我們有了一套拆穿性的譜系學，因為以下可能性依然存在：譜系學里提到的東西和道德信念所反映的道德屬性，是等同的或有依附關系的，而且正因如此，譜系學最終還是使得道德信念為真。（我在術語上假定，兩個東西有同一或依附關系不等于它們之間有預設關系）。在下文中，我將展開解釋這一點，并圍繞吉爾伯特·哈曼（Gilbert Harman）的著名論證來組織討論。本章有相當的部分是用于回應這種還原主義理論的可能性，說明我為什么覺得它站不住腳。

## 6.2　哈曼的挑戰

哈曼（1977，1986）詳細地討論了以下想法：因為我們不需訴諸道德判斷為真（即不借助這些判斷所反映的道德事實）就能夠徹底解釋道德判斷，所以從認識論的角度看，道德判斷就被削弱了。哈曼關注的不只是我們是否可以不訴諸道德判斷之為真來解釋道德判斷，更是我們是否需要預設道德事實來解釋世上的任何一件東西。很多人說哈曼的答案是否定性的（即他斷定不存在道德事實），但其實他的立場要更微妙一些。哈曼的結論其實是條件性的：如果沒有對于道德事實和自然主義事實之間聯系的還原解釋，那么道德言論就不能得到檢驗，道德理論就不能被證實或證偽，而我們也沒有道德事實存在的證據。但哈曼從來沒有斷定這個結論的前提條件成立。他沒有論證說從道德到非道德的還原不存在，他只是說這種還原不容易得到。事實上，他自己就提出了一個在他看來是合理的還原解釋，也就是說，他認為道德事實是有可能存在的，（想必）還認為我們的道德判斷可能是證成的。其實他還明確指出“（相對性的）道德事實的存在是有經驗證據的”（1977：132）。因此，哈曼不是個道德虛無主義者。他關心的主要是指出“這里有個真正的問題”（1986：67），指出“如果道德言論不能被還原成自然主義的言論，那么如何檢驗道德言論就是個真正的問題”（1986：59），并挑戰那些表述觀點過于粗率的自然主義者。

讓我們采用哈曼自己的例子入手。他讓我們設想“你看到幾個孩子給一只貓身上澆汽油，然后將其點燃”（1977：4）[[101]](#_101_51)。過了三頁，他寫道：

……你未經有意識的思考就立即做出了道德判斷，比如這些孩子點火燒貓是錯誤的……為了解釋你所做的（這個判斷），大概可以合理地假設那些孩子確實把汽油澆到了貓身上，你也看到他們那樣做的過程。但是（沒有）明顯的理由要假設所謂的“道德事實”，要假設放火燒貓確實是錯的……其實，對于你何以做出這種判斷的解釋，似乎跟有關道德事實的假定完全無關。我們似乎只需要假設你有一些還比較具體的道德原則，你出于道德敏感而做出的判斷反映了這些原則。對于我們的解釋來說，你當下的直覺判斷是真是假似乎根本不相干。（7）

哈曼沒怎么展開說明“道德敏感”如何足以解釋人們在某些情況下會做出道德判斷。哈曼的論證沒有必要涉及生物自然選擇；它只需要我們對于道德判斷可以給出一種完整的解釋，并且該解釋同道德判斷的真假無關。演化的解釋只是這類假說的一種形式，或者只是這類假說的一個可能的元素。有的人可以相信道德完全是文化的產物，因此只用文化的社會化來解釋所有的道德信念，而這一解釋并不預設這種社會化是學會道德真理的過程。（馬克思主義式的和弗洛伊德式的進路也許屬于這種觀點。）換言之，假如本書關于道德演化的說法都是錯的，哈曼的論證依然站得住腳。

哈曼的論證所引發的辯論，有些集中于探討道德經驗是否確實類似于感覺經驗：我們確實就是“看到”了行為的錯誤嗎？但是，對于我們來說這就跑題了，可以不必理會；即使道德判斷是由想象促發的，哈曼的主要觀點依舊成立。（我們很少有人親眼見過貓被燒死或者甚至被棍子打，但是我們都能夠判斷這種情景在道德上令人厭惡的。）比這一點重要得多的是，哈曼最終的立場并不是認為道德判斷對于解釋我們的道德判斷沒有作用（雖然他在以上引文的有些地方語氣不像）。他并沒有說它們是多余的，只是說如果我們在自然主義的解釋里找不到容納它們的位置，那它們恐怕就是多余的。他的論辯是一種挑戰，而不是表達勝利。人們通常認為，哈曼的挑戰代表了自然主義面臨的問題，但實際上要是自然主義者能合理地闡述她的理論，挑戰也就煙消云散了（需要記住，哈曼自己也青睞這一結果）。

我對哈曼的想法理解如下。設想一下，簡判斷說某個燒貓的行為是錯誤的。假設我們可以運用物理學和化學，既徹底又準確地解釋簡做的判斷，而在這套解釋中，“貓”、“燒”和“錯誤”甚至都沒有出現。這證明了燒貓在對于解釋簡的判斷沒有作用了嗎？沒有，因為燃燒的貓可以被還原為物理學和化學現象，我們由此也可以知道，在我們對簡的判斷的因果解釋中，燃燒的貓是（不妨這么說）暗中存在的。但是，至于錯誤這種性質怎樣被納入這個自然主義的世界，我們能給出類似的說明嗎？哈曼想指出的是，如果我們能對為什么在簡看來該行為是錯的給出一套自然主義的解釋，但又不知道錯誤這種性質本身如何能夠容納于（還原為）這個解釋，那么，這套解釋就完全不需要預設錯誤這種性質是現實存在的，所以我們就沒有理由相信錯誤的現實存在是燒貓事件的成分之一。顯然，這一推理完全是可以一般化的。每當我們判斷某件事在道德上是錯誤的，對于該判斷都要有一個完整的解釋，該解釋既不預設道德事實，也不是道德事實的還原基礎。既然我們無須借助道德事實就解釋了道德判斷，而且也沒有什么需要假定道德事實才能解釋的東西，（我們之所以想要對一種現象做出道德解釋，最終不都是因為我們做出了道德判斷嗎？）我們就沒有理由相信有任何東西在道德上是錯誤的（這一點也適用于其他道德性質）。

至于究竟怎樣算是適當的“還原”，哈曼講得比較含糊。他說還原“不需要包含嚴格的定義”，他還用桌子和一組原子之間的關系作為例子（1985：33）。這同上文說的把燃燒的貓“還原成”物理和化學現象的意思是一致的，即便在包含“貓”和“燒”的命題和用物理學和化學詞匯表達的命題之間，不存在可以相互推論的語義關系。然而，至少我們自認為可以理解燃燒的貓是如何融入物理學和化學描述的世界（我們可以解釋一只燃燒的貓是怎樣的物理學和化學物體）。這種還原的觀念是非常寬泛的，甚至可以包容許多一般歸于非還原主義的立場。在這場爭論中，哈曼著名的論敵尼古拉斯·斯特金（Nicholas Sturgeon）使用的還原觀念則狹隘得多，他認為“是否可以用非道德的詞匯來給道德解釋進行還原定義”（1985：59；同時參見Sturgeon 1986）。（在我看來，遺憾的是哈曼和斯特金的爭論中有太多篇幅都是源于這種概念差別。）接下來我將遵從哈曼的寬泛用法，因為這一用法能讓哈曼的論證更有道理。歸根結底，重要的問題在于：雖然道德判斷的譜系學是用非道德的語言表達的，但它是否在某種意義上還暗中“包含”了道德事實的存在。沒有理由要求道德事實可以用自然科學的語言來描述；純粹的本體論上的關系就足夠了。就連那些常被視作“非還原主義的道德自然主義者”的人都會同意，道德屬性的每個具體實例同時也是一組自然主義屬性的具體實例。這看起來就足以使他們在寬泛意義上成為哈曼一樣的還原主義者[[102]](#_102_51)。

圖6.1描繪了關于如何解釋道德判斷的現象的兩個對立的假說。需要記住，假說A應該是在經驗上得到證實了。忽視這一點可能會讓人覺得這個論證同標準的哲學懷疑主義挑戰大同小異，一點也不有趣。例如，也許你相信現在你正坐著讀書，但煩人的懷疑主義者會提出另外一個假說（“假說A”），該假說同你能得到的證據完全一致，但又宣稱你其實并不是在讀書。也許像笛卡爾想象的那樣，你是被一個全能的魔鬼蒙騙了（他以此打發時間），或者你只是一個大腦，漂浮在裝著營養液的缸里，神經受到的外來刺激讓你覺得你是在坐著讀書。要證明你能夠意識到這些假說是錯誤的，是出了名的困難，因為根據假設，在正常的假說和這些怪異的假說里，你的感覺經驗證據是一樣的。但我們現在討論的觀點不能這么輕易得出。現在的問題，不僅僅是我們可以編造的一套融貫假說，該假說指出我們的許多日常信念都是錯的；而是我們可以有經驗證據來支持這個假說，該假說可以解釋為什么我們有那些信念，但不需要預設那些信念是真的。這個論證不依靠認識上的證成的一些極端標準；這里的懷疑主義者不是讓人們設想一些通常會反遭嘲笑的那種缸中之腦式的可能性。如果道德證成的日常標準能夠采納有關人類演化的證據，而且，要是這些證據最后證明了道德信念是未得證成的，那也是按照日常的標準未得證成。

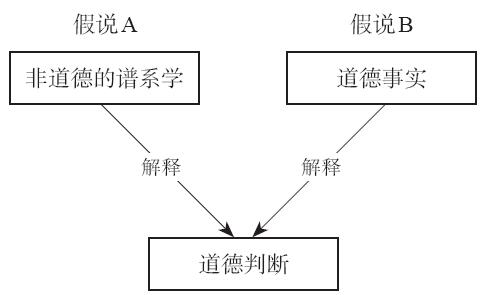


圖6.1

在圖6.1的基礎上，也許有人會覺得奧卡姆剃刀（Ockam’s Razor）應該可以迅速地把假說B從圖示中剔除出去，因為對道德判斷我們已經有了完整的解釋，在本體論上就不再需要預設道德事實這種額外的東西。邁克爾·魯斯宣稱“客觀的道德基礎是贅余的”（1986：254；同時參見Ruse and Wilson 1986：186—187），以此得出結論說道德演化的基礎削弱了道德，他似乎就是這么論證的。但圖6.2表明了這一想法太過草率。那條彎彎曲曲的線代表還原關系（在寬泛的、本體論的意義上）。如果道德事實可以還原為譜系學解釋中涉及的非道德事實，那么前者就不能夠出于簡省的考慮而消除，就好像不能因為我們能用物理學解釋貓，于是就可以把貓從本體論中消除。在某種意義上，貓是位于物理學和化學之上的一個本體論范疇，但在另一種意義上，預設這個范疇并不會令世界在本體論上更為繁富，因為這個生物范疇可以容納于更基礎的范疇。這正是說它可以被還原的原因。

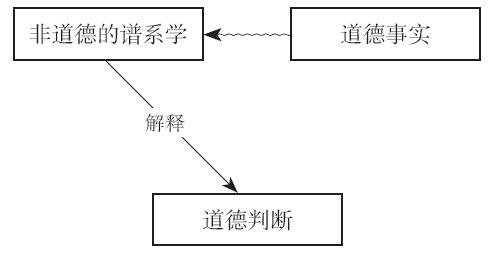


圖6.2

哈曼正確地堅持認為，不輔以一個具體的理論，光是承認道德自然主義就有可能解決這些問題，并用多少還算精確的方式解釋道德如何容納于自然世界，那基本上于事無補。就好像也許有人會說，要是能證明幽靈的性質可以妥帖地融于自然主義的世界觀，那么有關幽靈的說法也可以得到支持。確實如此。然而，這種情況本身絲毫不能提供相信幽靈的理由。除非見識到了某個還原解釋，并且以其為合理，否則全局的自然主義者不會愿意承認幽靈的存在。前面我在詮釋哈曼的意圖時，說他指責那些“表述觀點太過粗率”的道德自然主義，意思正是如此。光是保證存在著這樣的自然主義理論還是不夠的，除非我們確實有理由相信可以得到這樣的理論。另外，光是給出一個自然主義的還原也是不夠的，因為這輕而易舉。查爾斯·斯蒂文森（Charles Stevenson 1937：14）曾經說過：“一個東西是道德的，當且僅當它顏色粉紅，還有黃邊。”（他這是在提出一個哲學問題，而不是提出一個真的可能的理論！）論證的責任在道德自然主義者這一方，他們需要提出一套理論（或者至少給出讓我們相信可以證明這個理論的有力理由），除此之外，哈曼補充道，“我們還要能信服這套解釋”（1986：63）。在這個任務完成之前，一個得到證實的、非道德的譜系學似乎讓我們沒有理由相信我們的任何道德信念是真的。

這樣的還原解釋可能是真的嗎？道德自然主義的前景如何？我們在第五章看到了，摩爾那個想要畢其功于一役的反對論證失敗了。也許我們力所能及的只是逐一考察各個道德還原主義理論，就像我們在第五章考察規范演化倫理學的不同版本一樣。這樣的任務超出了本書的范圍。事實上，我認為（無論是還原還是非還原形式的）道德自然主義都面臨著一些普遍的問題，但在此我只會討論其中之一：這些自然主義都無法體現道德言論富含的那種不可避免的實踐權威性。下面的兩節將主要討論這一實質性的問題，同時我也將討論哈曼本人偏愛的那種道德的自然主義還原。

## 6.3　道德自然主義的前景[[103]](#_103_51)

相對于第五章的任務，我現在面臨的任務在某種意義上更為局限，但在另一種意義上也更為寬泛。在那一章，我討論了一個一般性主張，即演化理論可以在某種意義上維護道德；現在我討論的將是一種具體的維護方式：確立道德自然主義。另一方面，我在第五章也討論了“演化的道德自然主義”這種具體的立場，這種立場認為道德事實可以還原成本質是演化的和歷史性的事實。現在我也將考慮更一般意義上的道德自然主義，它只需要道德事實能夠還原成道德信念的譜系學包含（或許是偶然包含）的東西。要是這種自然主義是可行的，那么，雖然在表面上我們不須提及或預設道德事實就能解釋人類的道德判斷，但這其實就沒有拆穿的效力。我在此將概述在我看來道德自然主義者所面臨的最大挑戰，我認為也是無法克服的挑戰。

為什么斯蒂文森的玩笑式道德自然主義（“一個東西是道德的，當且僅當它顏色粉紅，還有黃邊。”）是不可接受的呢？這顯然是因為在我們看來，道德上的善應該是什么樣子，有一些必要的條件，而這些條件斯蒂文森的玩笑提議滿足的極少（也許完全沒有）。這些條件的總體給任何道德自然主義都加上了限制因素。我們可以否決任何極不符合這些條件規矩的自然主義說法；無論這種理論找到的是什么屬性，那都不是道德上的善。這不等于說一定要有“完美的匹配”。如果我們認識到了（比如）十五項要求，而找到一個滿足其中十二項的自然主義理論，同時又沒有其他更吻合的自然主義理論，那么可能這就足以讓我們信服了。（參見Lewis 1970，1989；Smith 1994，第二章；Joyce 2001a，第一章）但是光看數量，或者認為可以有某種決定怎樣才算足夠的普遍的“黃金比例”，都是不對的。也許對于有些概念而言，我們認為某一規矩是關鍵核心，理論只要不滿足這一條標準就可以將其否決。也許對于有的概念而言，我們愿意采取更為寬松的標準。

在第五章的末尾，我確定了一項這樣的要求，它對任何演化自然主義來說應該都很棘手。這個規矩跟我們怎樣找到（或者更確切地說，我們不是怎樣找到）道德真理有關。演化的道德自然主義意味著，我們認識演化真理的方式可以構成認識一般的道德真理的方式。但是，我論證道，這是個瘋狂的結論。我們不是（也不能）通過考古發掘，揭示人類演化的事實，進而發現埃爾金大理石是否應該歸還給希臘的。然而，如果我們考慮另一種形式的道德自然主義（比如功利主義），那么主張我們可以通過揭示有關如何提高總的凈幸福量的事實來發現道德真理，就顯得沒有那么愚蠢。我認為任何堅持這個主張的人都錯了（原因將于稍后揭曉），但錯的不算很離譜，因此也算不上是反證了功利主義的謬誤。因此，第五章末尾的論證對某種道德自然主義的破壞有限。那么，我們可以做得更好，確定一項任何道德自然主義都難以滿足的要求嗎？我相信可以。我在5.7節對凱斯比爾的演化道德自然主義所提出的批評可以進一步推廣，給所有的道德自然主義制造麻煩。

很難看出自然主義的事實如何能夠提供不可避免的權威，而后者顯然是我們希望和要求道德價值所具備的性質。在第二章，我討論了“道德不可避免性”和“道德權威”這兩個術語，還把它們的結合取稱為“實踐影響力”。讓我來再試著解釋一下為什么滿足這兩個條件是必須的。首先是不可避免性。我在2.4節（以及5.7節）詳細地論證了，人們不能求助于特別的欲望（“我就是喜歡殺害無辜！”）來避開道德禁令（“不要殺害無辜！”），也不能通過宣稱對道德價值沒有興趣來擺脫道德上的考量。我們認為，不論人們的欲望和興趣如何，道德價值和命令通常都是對人們具有約束力的實際考慮因素（雖然也許是可以推翻的）。關于這一點，我認同麥基的說法：一個人做出道德判斷，是想說一些“不只是純粹描述性的東西，更顯然不是對人毫無影響的東西，而是包含著對于行為或不作為的要求。這種要求是絕對的，不依賴于他自己或任何其他人的欲望、偏好、策略或選擇。”（1977：33）普通人大概不會說得這么簡潔和準確；我確信普通人對于道德的觀點還是比較不成形的。但是我們不應該把不成形和可以商榷（negotiability）混為一談：對于道德的不可避免性，普通人的看法可以是堅定不移的，雖然他們不能闡明不可避免性究竟是什么。當像麥基這樣的哲學家站出來，嘗試著精確表達和闡釋這種日常概念，以便評估其合理性，但我們不能因為他是從未成形的原料中打造出明確的理論，不能因為他們用的可能都是普通人不熟悉的詞匯，就得出結論說他們只不過是把自己哲學或文化上的偏見強加于普通人。當然，他有可能是這樣做的，但要確定是否果真如此，還要看他的闡釋是否成功地抓住了人們（他想要描述的正是這些人的概念）的日常道德實踐。在我看來，麥基的論據很充分。道德規定的使用很明顯是絕對的：我們會毫不猶豫地把它們應用的某些人身上，即便我們知道遵守道德不利于滿足他們那些離經叛道的欲望。假設當連環殺手杰克·特拉威克宣稱“知道我會面臨死刑……我還是會再犯一遍。”時，他是非常真誠的。承認這一點絲毫不會讓我們想撤回道德指責。我之前（在4.5節）也提到了，把某些違規行為視作獨立于任何權威的傾向，和把這種違規同謹慎務實（prudence）的規范區分開來的傾向，都是跨越文化的，而且在很小的孩子身上就已經體現。

許多道德哲學家認為，道德規定不僅是不可避免的，還經常（或者總是）充滿了某種權威。不妨同禮儀做個比較。雖然禮儀的規則也是不可避免的（不在乎禮儀的人也會違反禮儀的要求），我們一般都不認為禮儀制度對人有真正的約束力。一個人出于正當的理由而舉止粗魯（比如為了避免朋友吃下黃蜂，嘴里塞滿東西的時候還張口說話），仍是違反了禮儀，但他做的是正確的，而且事實上他沒有理由不這樣做。只有當遵循禮儀符合人們的利益時（我們多數人在大多數時候確實如此），才有真正的理由去遵循。人們一般認為道德需要更強有力和更具權威的東西。菲利帕·福特曾經說過，許多道德哲學家追隨康德主義的傳統想法，認為道德規定“必然給任何人都提供了行為的理由”（1972：309）。一條規定可能在某人沒有任何理由遵循的情況下，不可避免地適用于某人；但如果這條規定同時也（可以說是“自動地”）提供了遵循的理由，那我們就說它對被規定的人而言具有權威。

上文著重強調了道德的實踐影響力，在此基礎上，我們可以把道德自然主義者歸入兩個陣營：（i）那些試圖證明自然主義的框架可以容納這種實踐影響力的道德自然主義者，（ii）那些根本否定道德有這種要求的道德自然主義者。在后一陣營里，有的論者假定所謂不可避免的道德權威只不過是哲學家的幻想（他們特別反對道德的權威性），而其他人則同意人們通常的確覺得這一影響力對于道德很重要（是成為道德所需的必要條件）——甚至或許還接受“這個假設是道德詞匯基本的、約定俗成的含義的一部分”[[104]](#_104_51)——但與此同時，他們又認為符合了其他足夠多的圍繞著道德的陳規，就差不多可以算是合格的自然化。我將在本節接下來的部分討論（i）；6.4節將回頭討論（ii）。這兩種自然主義立場的根基在我看來都不牢靠，而這就有力地支持了哈曼的挑戰，最終也支持了“我們的道德判斷是未得證成的”的觀點。

一個道德自然主義者要想解釋不可避免的實踐權威性，就必須在自然世界里面為這種權威性找到合適的位置。讓我們假設，這就等于要找到行為的某種理由。日常話語賦予了道德某種比禮儀更大的權威性（該道德自然主義者是這么認為的，而我也傾向于贊同），不只是因為人們覺得道德比禮儀更為重要；而是因為道德對人的“約束”應該要更直接有力。引入理由的概念，可以看成是哲學家在嘗試把日常言論和實踐里面模糊卻重要的東西塑造成一個準確的觀念；但這也把問題復雜化了，因為圍繞著什么是理由有許多混淆和爭論。人們甚至都可以說起“禮儀的理由”，依照這種說法，禮儀的規則確實給出了理由；但這種理由純粹只是制度建構的產物而已（其意思無非是：只要禮儀規則規定了X做ϕ是證成的，禮儀制度就允許我們說“X有理由做ϕ”）。對于那些被賦予理由的人，允許這樣的說法并不能構成真正的實踐考慮因素。道德自然主義者想把實踐權威性安置于世界之中，只找到某種包含以下規則的人類制度是不能令他滿足的：在某些情況下，可以宣稱“X真的應當做ϕ”，“X有理由去做ϕ”，“無論她興趣如何，X都有理由做ϕ”，甚至“有個獨立于心靈的事實，要求X做ϕ”。他不會滿足于只是找到一個制度性的規范框架，該框架包含著什么行為是證成的規則的，因而也能產生“證成理由”。（關于證成理由和動機理由的區分，參見Smith 1994：94ff）。就像尋找塔斯馬尼亞虎[[105]](#_105_51)的人不會滿足于找到一群相信塔斯馬尼亞虎還存在的人，這種道德自然主義者想找到的是真正的不可避免的實踐權威，而不是一個允許言語之間仿佛這種權威確實存在的人類制度。

許多道德自然主義者在道德和理由之間引入了一定的中介。比如功利主義的一種策略，就是先說道德上的善等同于幸福，然后再試圖證明我們有理由追求普遍的幸福。一種更簡單的策略則宣稱，道德要求本來就是我們有真正理由去遵循的事情（假設對于怎樣算真正的實踐考慮因素的觀點本身也可以自然化）。后一種進路是哈曼表示贊許的道德自然主義版本，他稱之為“實踐理性理論”[[106]](#_106_51)：

這種觀點認為，一個行為是錯的就意味著行為者有充分的理由不那樣做。這就等于是說，如果行為者進行了正確的推理，那他最終就會決定不那樣做。

依照這種觀點，對于簡那有關阿爾伯特（他殺害了那只貓）的（道德）信念，我們可以用阿爾伯特行為的實際錯誤來進行部分的解釋。阿爾伯特的行為是錯的，因為他有充分的理由不那樣做。如果阿爾伯特進行了正確的推理，他最終就會決定不那樣做。如果簡了解了相關的推理，而且這就是她為什么相信阿爾伯特的行為是錯的，那么她的信念就從阿爾伯特行為的錯誤中得到了部分的解釋。

在哈曼（以及許多同意他的人）看來，一個人“有充分理由”做什么，跟假如他“進行了正確的推理”就會想做什么是密切相關的。對此，一個自然的問題就是：“這為什么就有了實踐權威呢？”假設阿爾伯特的某一個可以選擇的行為，具有假如阿爾伯特進行了正確地推理就會想去做的屬性，為什么對于阿爾伯特而言，這個屬性就代表了真正的慎思因素呢？我認為答案就在于他不能合理地懷疑這種屬性對他而言是不是實踐考慮因素，因為這種真誠提出質疑的行為本身就表示已經部分采納了正確推理的標準。而這只不過是懷疑和提問這些行為所具有的一般特質。提出“為什么要X？”的問題，這一行為本身就預設了提問人是在尋求答案，也可能會接受某些答案；預設了提問人同意被提問的一方也許在認識上處于比自己更為優越的地位。但是承認自己的認識可能不足，也同意進步值得向往，就是認同那些你要是正確推理就能得到的意見確實帶有慎思的力量。因此，“我承認如果我推理正確就會想要ϕ，但這跟我有什么關系呢？”這種懷疑是說不通的。

小結一下：我們現在考慮的問題是道德是否能還原成自然事物。如果不能的話，圖6.2就不可接受，而圖6.1似乎就是正確的，于是奧卡姆的剃刀就使得我們沒有理由相信道德事實的存在。為了評估任何有可能成立的還原理論，我們都要檢查它們重點強調的自然性質是否能夠滿足我們賦予道德價值和規定的前理論標準。一個可能的標準是道德通常具有的、不可避免的實踐權威，所以有一類道德自然主義者試圖在自然世界里找出實踐影響力（另一類自然主義者則認為這根本不是一條適當的標準，后文將探討他們的觀點）。這些自然主義也許想在實踐權威和理由之間建立直接或間接的聯系。但是理由有很多種，所以有必要消除歧義。我們想找到的是真正的慎思因素（即哈曼說的“充分理由”），而不只是某種制度所賦予的因素。更準確地說，我們想找的是對一個人要是推理正確（即要是完全理性）的話就會想要怎么做的一套解釋[[107]](#_107_51)。因為這里的企圖是把道德自然化，這里的“正確推理”概念無論是什么，都要能夠以自然主義的方式理解。例如，我們不能將其理解為遵循道德德性的推理或依照對自身道德責任的感知而進行的推理，否則就把整項工作帶入了惡性的邏輯循環。

但這里的問題，同時也是我懷疑這項計劃的主要原因，就在于要是想把道德影響力自然化，我們就不能只找到具有實踐權威的屬性（我們大概已經在你要是推理正確就會想做身上找到了這種屬性）。我們也必須滿足不可避免性；我們需要找到一種無論人們興趣如何都對其有權威的屬性。然而，自然主義能否實現這一點，很令人懷疑。假設厄尼恰好很享受喝紅酒但不喜歡啤酒；也許要是他推理正確，就會得出結論說每天晚上來一杯紅酒對他來說可以接受的，但要避免喝啤酒。再假設伯特不喜歡紅酒卻喜歡啤酒；要是他推理正確，也許就得出結論說每天晚上來一杯啤酒對他來說是可以接受的，但要避免喝紅酒。厄尼和伯特的推理都會是正確的，但兩者趨向不同。兩者趨向不同顯然是因為他們兩人本來的品位和欲望就不一樣，而正確的推理雖然未必接受所有欲望（比如厄尼的推理讓他控制住喝掉一整瓶紅酒的欲望），但還是會對這一點給予考慮。至少在這個例子里，一個人正確的推理得出的規定與他的興趣有關。（哈曼實際上也同意（1975：9）：“如果S說A（在道德上）應該做D，S的意思是A有理由做D，S自己也同意該理由。我將假定這些理由的根源是在S看來A具有的目的、欲望和意圖。”）

既然我們確信，在某些情況下，正確的推理會因人的偶然欲望的不同而不同，那么現在關鍵的問題就是“為什么我們不能認為情況一貫如此？”考慮一下哈曼筆下那個討厭貓的家伙，阿爾伯特。如果阿爾伯特本來就有邪惡的欲望（我們可以這么假設），我們有什么理由假定他通過“正確的推理”就能廢除那些欲望呢？推理確實經常讓我們意識到我們目前某些欲望是不合適的，所以也許阿爾伯特要是反思得更仔細，他最終就會覺得傷害貓不是什么好事；但也許他還是不會那么想。也許正確的推理只是讓阿爾伯特知道了怎么更有效地傷害那只貓。道德自然主義者需要的，顯然是對于“正確實踐推理”（或“實踐理性”）的一套實質性的、自然主義的解釋，而且根據這套解釋，無論他們各自本來的欲望是什么，每個人都可以通過這種推理得到一致的，并且還大體上符合我們眼中的道德要求的實踐結論。（后一個條件也是必要的，否則達成的一致結論可能是道德總要求我們把自己擺在最重要的位置，盡可能地把東西據為己有，只要能免于懲罰就欺騙他人等。）但是還沒有人能給出像樣的解釋。當然，有些人嘗試過，但給出的解釋都遠未得到一致贊同，并且都充滿困難。

現代哲學做出的，最重要的嘗試是大衛·科普所說的“自我觀念策略”（“self-conception strategy”）。他的描述如下：

根據自我觀念策略，一個頭腦清楚的理性的人必須以一種特定的方式設想自己，但是假如她不把道德理由當作權威，那她就不能前后一致地以這種方式設想自己。比如有人說，不把道德理由當作權威的人就不能把自己視為是自律的（autonomous）；或者說她就不能把自己看成或珍視為理性的反思性行動者（ref lective agent），是出于理由而行動；或者說她就得信奉實踐唯我論（practical solipsism）；就不能一致地期望她給別人提出的理由會得到回應；諸如此類[[108]](#_108_51)。（2004：35）

但是在我看來，光是指責阿爾伯特在傷害那只貓的時候也必然地削弱了他的自律，或指責他沒有把自己珍視為反思性的行動者，指責他信奉實踐唯我論，他損害了對于別人回應他自己的理由主張的期望，這些都很難有什么成果。即便這些責難成立，還是需要問“破壞自己的自律，能把自己珍視為反思性的行動者等，為什么就那么糟糕呢？”這些事情聽起來是有些令人嫌惡，但這“令人嫌惡”的到底是什么？克里斯汀·科爾斯嘉德（Christine Korsgaard）[[109]](#_109_51)寫道，在道德上違規就是違反了“對我們來說最為重要的自我觀念……是失去了你的完整性，因此也失去了你的身份……是無法再覺得你自己有價值，覺得你活得有意義……對于所有實踐目的來說，就都等于是死了或者比死還不如”（1996：102）這樣的評價，讓人覺得作惡的人是個心理病態的禽獸，只要能逃脫追究，為了自利情愿把親友趕盡殺絕。但是不要忘了，這些說法本來要揭示的是所有道德違規的錯誤之處：借書不還，對服務員無端地舉止粗魯，偷走一份賓館走廊上的報紙。這些行為大概都是錯的，但是說它們讓犯錯人“死了或者比死還不如”（或者哪怕只是接近這一不好的境地），實在是太過夸張，沒有根據。有的人可以偶爾有道德違規，但總的來說過著幸福滿足的生活，這一事實就應該讓我們懷疑這個人是否真的像自我觀念策略的支持者說的那樣破壞了自己的行動主體性（即便這一疑慮可以得到解答，我們還是想要知道為什么這種微小的自我損害就那么糟糕）。這些所謂的自我傷害，渺小而難以確定，這本身就提出了一個問題：為什么避免這些傷害是至關重要的？科爾斯嘉德認為，犯錯人違反了他作為一個人的身份，這是他的其他身份（父親、學者、共和黨人、罪犯等）的基礎。據說前一種身份更為“深刻”（同上：258），因為擺脫了這個身份也就不可避免地損害了其他身份。可是（正如我們在上文提及凱斯比爾的演化道德自然主義時了解到的）這種反事實的不對稱性為什么以及如何具有實踐重要性，這就不大明顯了。也許就像這個反事實句描述的那樣，我作為人的身份確實比我作為父親的身份更為深刻，但是似乎不能由此推論出作為人所產生的任何價值，比起作為父親所產生的價值都更重要，更強大，或者更有權威。

鑒于自我觀念策略一直以來都沒能對獨立于欲望的正確實踐推理給出一套合理的解釋，也鑒于我們知道有些情況下正確的實踐推理是會隨著欲望的不同而不同的，似乎我們就可以不失合理地假設（就算只是暫時性的），一般而言正確的實踐推理都是隨著欲望的不同而不同的，實踐理由相應的也是如此。這意思不是說一個人有理由做某件事就等同于當且僅當他想要那樣做。遠不是這樣。正確的推理可能揭示了我的某些欲望是錯誤的，愚蠢的，或者不適當的（因此我并不一定有理由追求滿足這些欲望）。正確的推理還可能讓我意識到應當追求某些我實際上并不欲求的東西（因此我并不一定對于我有理由去做的事情有欲望）。這里的要點在于，我有什么理由在某種程度上取決于我的實際欲望、興趣、計劃和目的，所以在相同的環境里，不同的人可以有非常不同的理由。（用哈曼的話說，一個人的理由“來源于”她的欲望。）

我花費了很大的工夫，探求道德自然主義者能否通過把不可避免的實踐權威和對于充分理由的一套在自然主義上可行的解釋聯系起來，從而在世界里找到這種實踐權威的容身之地。他們這條路行不通。這一結果并不意味著實踐影響力這個概念本身是不一致的，因為非自然主義的正確推理概念有可能符合實踐影響力的概念。從前文得出的結論或許是：如果實踐影響力確實是道德話語的重要環節，那么這種權威性要是能用一套對理由的解釋來闡明，這里的理由必須是非自然主義的。這顯然不利于我們的道德自然主義者。道德自然主義者的另一種選擇是徹底放棄理由的觀念。畢竟我也坦承了，這些對于理由的談論，大概最好是看成是哲學家在盡可能地解釋一個日常的觀念，而并不一定是普通人明確乃至暗中相信的；或許用理由來闡釋道德的權威是錯誤的。但是我不得不承認，我是不知道如何順著這種選擇開辟出一條路徑（哪怕是一條能隱約看到是死胡同的路徑）。什么樣的規則可以既對一個人有權威，但同時又承認他可以正確宣稱他沒有真正的理由遵循？一個人要是愿意承認道德惡人可以沒有理由遵循道德，他就不大可能還自認為是在為某種不涉及理由的不可避免的道德權威辯護。他更像是干脆直接放棄了滿足這個條件的嘗試。但這又有什么糟糕的呢？這就把我們引向了道德自然主義者的第二種策略：不再認為這個標準的一定要得到滿足，或者干脆否認它原本就是個標準或條件。

## 6.4　誰需要道德影響力？

論證現在到了關鍵的時刻。一方面，許多哲學家都認為，不可避免的實踐權威是任何合格的道德理論都要滿足的特質。我在上一節論證了如果他們的想法是正確的，那道德自然主義就有麻煩了。誠然，我只討論了一種類型的道德自然主義，即實踐推理理論，但我是仔細挑選了這個例子來表達一個更具一般性的觀點。這是因為，要是連把道德事實和有關理由的事實直接等同起來的道德自然主義都沒法完成任務，那些只是想把道德事實和實踐理由間接聯系起來的理論還有什么希望成功呢？因此，我的結論（但我不會裝作是蓋棺論定）是：道德自然主義不大能找到辦法來提供這種實踐權威。

另一方面，許多哲學家也否認這一特質是不可或缺的；他們會滿足于把某些自然的屬性稱為“道德上的善”，就算該屬性沒有這種特殊的實踐力量。也許有時候人們接受這后一種觀點正是因為意識到自然主義無法給道德權威留下位置。但這些自然主義者常常花大量篇幅夸大宣揚他們的理論所能提供的不論什么“弱一點的力量”，這一事實本身就表明具備某種直截有力的實踐力量（比如實踐影響力？）是道德理論要滿足的條件。無論如何，我覺得這后一種立場是不現實的，在我看來更合理的結論是道德自然主義確實遇到了麻煩。（需要記住，承認這一結論并不會削弱人們信奉的方法論上的全局自然主義；后者容許人們還是可以在如下意義上的“把道德自然化”：給道德體制和實踐、道德心理學、道德譜系學提出一套科學上可以接受的解釋。）

避開了道德影響力的道德自然主義者可以承認，他們選出的自然屬性不能確保人們總是有理由按照跟該屬性相聯系的規定而行動，但是他們也許覺得，這種屬性和人們的理由之間只要有可靠的偶然關系就足夠了。比如那個功利主義者，她也許不再會力圖證明在把幸福最大化的行為和我們有理由進行這樣的行為之間存在著必然的聯系，但是只要能指出我們確實經常關心總體的幸福，而且這就為促進幸福的動機和理由提供了基礎（雖然只是在多數情況下有的動機和理由），這就足以讓她滿意了（例如，參見Railton 1986；Brink 1989；43ff）。這個提議可能不足以給我們帶來所有我們本希望道德理論能提供的東西，但我們也許應該聽取那句格言“足夠接近就足夠好”。

但是這確實足夠好了嗎？我們沒有原則性的決定辦法。這場爭論的解決似乎要取決于某些東西是否屬于我們的道德概念的核心性質，但是哲學家們沒有一致的見解來裁決乃至用概念表達這種爭論（極為類似的觀點，參見Nichols 2004：192—193）。要是一個人覺得某些東西是某個概念必須的特征，但另一個人表示反對，他們怎么爭論下去呢？大衛·劉易斯采用了嚴格說來（speaking strictly）和粗略說來（speaking loosely）的區別［（1989）2000：93］；“嚴格說來，麥基是對的：真正的價值需要滿足一些不可能滿足的條件，所以認為這種價值存在是錯誤的。但粗略說來，這個名號也許可以封給某些不完全當得起它的候選者……怎么處理這種情況主要還是性情問題。”雖然在一定范圍內，這種說法是無可反駁的，但其實這只是推后了問題，因為我們還是可以繼續爭論某些東西是否存在，即便我們已經限定了自己只是在粗略地說。我們大概不會接受某個允許幽靈和女巫存在的理論（即便只是粗略說來），而且我們也肯定不愿意接受某個容許問題可以依照“性情”解決的理論。

缺乏明顯的答案導致這成為許多元倫理學爭論的終點。我自己的看法是：這種問題的答案取決于相關的語言使用者會做出怎樣的集體決定。有時候新發現會讓我們決定一個概念（比如燃素和女巫）是不可能的；有時候我們更傾向于修正原有概念，除去有毛病的成分，然后大致延續原來的使用（同時性的概念在發現了同時都是相對的之后繼續被使用，聚合物的概念在發現了聚合物是大分子而不是膠體之后繼續被使用）。但是，我們根本沒有理由假設人們在做這種決定時是遵循了某種潛在的規則。憑什么說我們在這種事情上的集體決定就不會被最瑣碎的事情影響呢，比如一首廣告歌或者流行電影里一個詞的用法？這也表明，在很多情況下，某些東西是否屬于一個概念的核心性質，根本不是個客觀事實的問題。也許在一種社會氛圍下我們會決定答案是肯定的，但在別的（但也不是天差地別）氛圍下我們又會決定答案是否定的。正因如此，這樣的畫面看起來很有幾分無望：一個形單影只的哲學家，自信地斷言某些有爭議的性質是或不是某個概念的先驗要求，或者宣稱一個不完全地滿足了某個概念的規定的東西“足以”或不“足以”算作那個有缺陷的原有概念的延續。這種斷言最多也只不過是一種假設性的預測，預測的是一群人在我們面臨的環境里要是非選不可的話會做出怎樣的選擇（這種預測顯然充滿困難而且經常無法得到實際驗證）。

不過，希望或許還沒有完全喪失。我們可能還是有機會得到答案，只是需要考慮一下那個概念是做什么用的，是為了支持什么實踐活動，然后再問：修正后的概念，去掉了有問題的成分，是否能夠繼續扮演過去的角色？例如，即便我們知道了事物都沒有絕對同時的，相對的同時概念還是可以順利地頂替上來，因為在日常環境里它一樣可以適用于那些運動速度遠不及光速的生物。我們可以像使用絕對的同時概念一樣使用相對的同時概念，這就表明這里的改變不是用一個不同的概念替換掉原先的概念，只不過是對同一個概念進行了修正。因此我們沒有必要接受極端的立場，認為在愛因斯坦以前所有宣稱兩個事件同時發生的斷言都是錯誤的。與此相反，當我們發現了宇宙里沒有超自然的邪惡力量，女巫的概念對我們就沒用了。也許我們還是可以繼續用“女巫”這個詞來指那些處于正式社會的邊緣的、在當地文化里起到某種作用的女性（也許我們還可以找到一系列這些女性所共有的，也只有她們才具有的一組自然屬性），但繼續使用“女巫”并不能讓我們達到過去使用這個詞的目的：譴責這些女性的邪惡影響，證成殺害她們的行為。因此，再繼續用“女巫”這個詞來指稱某些現實的自然屬性沒有什么意義；于是我們放棄了這個詞，得出結論說歷史上所有“某個女人是女巫”的說法（就算只是粗略說的）都是錯的。

所以我們要問的問題就是：假如道德規定和人們遵循的理由之間只有可靠的偶然聯系，那么道德話語還能不能繼續扮演它實際上扮演的角色，無論這種角色是什么。如果不行的話，我們就有理由懷疑這種框架根本就算不上是“道德”體系。

一種跟人們的理由只有偶然聯系的道德在某些很重要的方面也近似于禮儀制度。這意思并不是說道德很淺薄；這只是對于道德命令的邏輯地位所做的論斷。在道德上犯錯的人（特別是如果他們相信不會遭到懲罰），其正當性就和你我在自己家中違犯禮儀時所具有的正當性一樣。這里用“正當性”（“legitimacy”）這個詞出人意料，所以讓我來試著解釋一下。如果我獨自在家的時候，坐在電視機前像豬一樣開懷大吃，我所做的顯然從禮儀的角度看是不正當的。要是有人看到我的吃相，說我的行為粗魯惡心，這是再正確不過的；其實我自己大概都知道我的行為確實體現了這些性質，但在此場合我并不在意；在這種場合下我可能覺得粗魯是無所謂的，或者甚至還會從中得到愉悅。當然，要是在家里不禁表現粗魯會讓我在別的場合陷入麻煩，那就是另外一回事了；所以讓我們來假設情況并非如此（假設我在所有社交場合都禮貌得完美無缺，私下的粗魯對此毫無影響）。那么，我獨自一人時為什么就不應該像豬一樣吃飯呢？按照我們的假設，我沒有真正的理由不這么做（我們假設即使我進行了正確的推理我還是會贊同在此場合這么做）。我們只能從禮儀的角度做出譴責。從那個角度看，我的行為是不適當的。但這我知道。只是在這種情況下我選擇忽略這個特定的規范框架——而這看起來完全是正當的。從禮儀的角度確實不然，但是在某種更具一般性的、超越體制的意義上是正當的。鑒于我在此場合的欲望（尤其是我對禮儀體制暫時的無所謂態度），總的來說像豬一樣吃飯是我應當做的。

可以舉一個例子，雖然略顯愚蠢，但或許能讓這個結論更為明顯。假設我聲明：“我在此宣布不論何時，只要我們遇到了秋天在挪威制造的東西，就可以說‘人人都要追求這個東西。’”假設你對此提議回應道“同意”。我們現在就創造了一個小小的規范體制，要求每個人都應該追求挪威秋天的產品。這個體制宣稱這樣做是證成的，因此遵循了該體制也能夠導致相應的理由主張：每個人都有理由（是證成的理由）追求挪威秋天的產品。但是，顯然我們并沒能真的創造出任何在慎思上有分量的、人人都要聽從的考慮因素。也許從我們的角度看，任何對秋季的挪威產品無動于衷的人都應該受到指責，但從客觀的角度看，那只是牽強的譴責，而且要是理智健全的人聽說了這樣的體制，或這些理由主張，或受到這樣的批評，他們在自己的慎思中無視于此將會是非常正確的。

記住這些想法，讓我們再來想象一下，如果道德也跟人們的理由只有偶然的聯系，那道德會是什么樣子的。假設杰克是真的真的想要謀殺約翰。在漆黑無人的曠野里，他同約翰單獨待在一口廢棄深井的邊緣，心想“為什么不呢？”被抓住的概率實際為零。也許有人會說，這樣做了就會鼓勵杰克在其他不那么容易逃脫懲罰的場合謀殺別人，或者他此后一直會帶著負罪感。這樣的話杰克也許就有了不謀殺約翰的理由。但我們現在打交道的是一個道德自然主義者，他承認道德跟人們的理由之間的聯系是偶然的，這就讓我們可以設想杰克沒有這些住手的理由。因此禮儀的情況也適用于這里。為什么杰克不應該謀殺約翰？我們假設，如果他進行了正確的推理就會贊同在此場合謀殺約翰。我們只能從道德的角度做出譴責。從那個角度看，杰克的行為是不適當的。但這他知道。只是在這種情況下他選擇忽略這個特定的規范框架——而這看起來完全是正當的。從道德的角度看確實不然，但是在某種更具一般性的、超越體制的意義上是正當的。鑒于他在此場合的欲望（尤其是他對道德體制暫時的無所謂態度），總的來說謀殺約翰是杰克應該做的。

我認為，這就讓道德顯得非常怪異。我們一般不大會先說杰克的行為是邪惡的和道德上不可接受的，然后緊接著又說他沒有理由住手，說其實總的來說謀殺是他應該做的。這不否認我們可以跟之前一樣繼續唾棄杰克的罪惡，也不否認要是抓住了他我們也有理由去懲罰他。（假定杰克在很久以前就殺了約翰，然后高興地去拜訪他的墳墓，也沒有遭到懲罰；直到最近我們才發現了可憐的約翰的尸體，以及他臨死前草草記下的罪行過程。）問題是，一旦同時坦誠了杰克的惡行可能有充足的理由，這些判斷就失去了部分的力量。也許這一坦誠讓人不自在是因為我們對道德體制的堅守已然根深蒂固，不愿意踏出它那些成文的規則。但看上去這本身就是有問題的，因為這似乎意味著堅持道德的人在道德規則面前有一定程度的自欺。如果根據假設，杰克確實沒有真正的理由住手，但有道德的人卻因為堅守道德體制而不愿承認，那么這種道德體制就在某些方面蒙蔽了他們。這就好像一個狂熱地堅持禮貌的人無法意識到，有些情況下禮儀可能在人們的慎思中不應該占什么分量。如果這種版本的道德自然主義（“不含影響力的自然主義”）是正確的，而且也允許我們承認它是正確的（即如果它顯然是正確的），它應該也允許我們說“從道德的角度看，傷害約翰是不可接受的”。對于杰克行為的罪惡，我們還可以補充道“但杰克在這個場合有充分的理由做出惡行，也沒有真正的理由住手”。

另外，在我們自己的實踐慎思中，面臨的道德決策遠不像殺不殺人那么重大，也沒有什么東西能阻止我們誠懇地補充上這種限定。想象在某個清晨，你站在空蕩蕩的賓館走廊上，想要從陌生人的門外拿走報紙，因為你自己點早餐的時候忘了要。假設你知道這個行為具有在道德自然主義者看來等同于道德惡的那種自然屬性。因而你知道這個行為在道德上是錯的；你知道從道德的角度看那是不可接受的行為。但現在考慮的這種道德自然主義仍然允許你進一步發問“我有什么理由要在意這里的道德錯誤呢？”（我覺得這種情況很奇怪，但那些自然主義者確實是這么說的。）誠然，是有一些道德上的理由要求在意道德上的惡（也就是說，有一些關于從這一特定規范框架的角度看來某些事情得到了證成的理由）。但這并不是你想找尋的那種理由，因為你問的其實是，在這種場合，對于這個行為，為什么要在意那個證成框架。注意到有道德得到證成的理由不拿報紙，只會讓你疑惑有什么理由要在意那些理由。

一個人不禁想要偷報紙，問自己“我有什么理由要在意這一件事的道德錯誤呢？”讓我們設想在這種情況下可能有三種回應（為了簡潔，這些回應將會是理想化了的，但它們是用來囊括各種可能性的）。首先，她可能意識到她完全沒有理由。她就像是上一個例子里的杰克，因為她的欲望，所以對她而言理性的做法是偷報紙，同時她也可以從容地對自己坦誠這么做在道德上是錯誤的。除此之外，她也可能得到肯定的答復，也就是說，這個道德錯誤可能跟她的欲望有一定的聯系，使得她確實有理由克制。這里的欲望可弱可強，導致的理由也可微弱可有力。我們可以根據后一點把弱欲望（產生脆弱的理由）和強欲望（產生有力的理由）作為一個連續可能性的兩極，分別代表第二和第三種回應。于是，一個人可以得到的第二種回應是意識到自己有理由避免犯道德錯誤，但理由只是微弱的；換言之，道德上錯誤的做法會讓她的某些欲望無法得到滿足，但這些欲望是可以讓步的，也是比較膚淺的，從而只要對她的欲望稍作改動，就能讓她避免道德錯誤的理由煙消云散。第三種回應則是具有遵從道德善的、堅定而根深蒂固的理由，因此也就有強有力理由避免道德錯誤。確實，這個人的欲望可能只是偶然的，但偶然不能和脆弱混作一談。（我關懷我在襁褓中的孩子的欲望只是偶然的——我不會因失去了這個欲望而不再是我自己或不再是人類——但也是牢固可靠的。）

這三種可能性各有各的問題，因而蘊涵著這三種可能性空間的那種不包含影響力的道德自然主義也是有問題的。我已經著重強調了為什么難以接受第一種回應，后者認為有的人也許有充分的理由去犯道德錯誤（即便同時也把他們要做的視為“道德錯誤”）。讓我再展開評論一下另外兩種可能性。

第二種人令人有些不安，因為對她而言，偷的誘惑本身就代表了一種促使她修改那些更愿避免錯行的欲望的理由，從而促使她再沒有理由不去偷報紙。換句話說，要是能丟棄那些要求此時遵循道德的欲望，同時又不引發巨大的心理混亂，那么不道德行為所滿足的欲望本身（即輕易獲得報紙的欲望）就會鼓勵她當場就在賓館走廊上改變她那傾向道德的欲望。這并不要求她放棄對于道德一般帶來的利益的興趣；相反，她只是調整了她的欲望，以容許偶爾的例外。（欲望不總是這么靈活和易受控制，但通常是這樣。）她不是在冷嘲熱諷“道德，虛偽的道德！”（因為一個其自身欲望完全沒有給她帶來在意道德的人，很可能是精神病態，但我這里說的只是普通人的弱點），相反，意識到一個理性的人可以有理由也可以沒有理由做出道德惡行，這會促使她心想“嗯，一般而言道德當然是好的，而且我總的來說當然有理由遵守道德，但我只是在當下為了點微不足道的小事而違犯了道德要求，并不會遭到哪怕一點點不利的結果”。一旦放棄了道德權威性，一旦道德的實踐效力取決于行為者偶然的欲望，并且行為者也知道事實如此，那么一個令人擔憂的后果就是人們在任何場合都可以在一定程度上主動決定自己有沒有理由做出符合道德的事：為了避開任何道德要求，人們都可以動搖自己的欲望（不道德的回報要是足夠豐厚，還可以徹底改變欲望）。不道德的誘惑本身就鼓勵人們這樣做。

第三種類型的人有強烈而堅定的欲望來抵制這種錯誤的行為，由此產生了堅固而重要的理由來遵循道德。道德自然主義者想讓我們覺得大多數人都屬于這一類。他們可能是對的，但是在知道在道德自然主義者看來的“道德錯誤”（以及“道德正確”等）究竟指的哪一種（或者哪一組）在自然屬性之前，我們還很難完全判定對錯。我當然同意我們大多數人，在大多數情況下，確實有堅定的理由避免做出一般被認為是“道德錯誤”的行為：有好的理由不偷竊，避免違背諾言，避免施暴等等。道德自然主義把道德屬性等同于某些自然主義屬性，他們這樣做的同時顯然也想支持廣泛接受的道德觀念（無論他們把道德錯誤等同于什么自然屬性，這種自然屬性都是偷竊行為所具有的，無論他們把道德正確等同于什么自然屬性，這種自然屬性都是遵守諾言所具有的，諸如此類），因此我同意他們很可能支持一種理論，該理論認為我們大多數人在正常情況下都有可靠而有力的理由做道德的事情。比如，假設等同于道德錯誤的自然屬性就是造成不必要的痛苦。這和我們的許多直觀印象是一致的。我認為我們大多數人確實有很好的理由避免造成不必要的痛苦；因此，要是偷走賓館走廊上的報紙會導致某些不必要的（雖然大概只是輕微的）痛苦，對于這種偷竊行為，我們的態度就會跟這第三類人一樣。

但是，無論這第三類人的決斷有多么堅定，他們對于道德的態度還是有些奇怪（至少是在不含影響力的道德自然主義得到支持的情況下）。問題在于，對于這類人的理由和動機而言，使用道德概念所做的思考似乎完全是多余的。如果她想知道為什么她不應該偷報紙，只要意識到那樣做會造成不必要的痛苦，她就可以得出答案。把那種有因果效力的屬性等同于道德錯誤似乎對她的理由和動機毫無增益。（把“造成不必要的痛苦”換成其他不含影響力的自然屬性，結果也是一樣。）其實既然道德錯誤就是這種有因果效力的屬性，將這個行為視作道德錯誤怎么可能有所增益？如果聲稱或認為這個行為是“道德錯誤”增添了某些實質性的東西，某些單憑聲稱或認為該行為具有沒有自然性質無法得到的東西，這就構成了反對道德自然主義者的理論的證據。因此我們給這個道德自然主義者提出的問題是：“按照你的理論，為什么我們還要一套單獨的道德話語呢？”為什么不開誠布公地說我們喜歡什么，不喜歡什么，什么有助于社會和諧，什么不利于社會和諧，什么東西我們能容忍，什么東西我們會懲罰？（我想起了鮑勃·迪倫寫的上帝和亞伯拉罕的對話：“你想做什么就去做吧，亞伯，但下次你看見我來了最好趕緊逃跑。”[[110]](#_110_47)）

對于“我們為什么需要一套單獨的道德話語？”的問題，我在本書前面的部分給出了一個答案（參見4.2節）。我提議說，相比非道德化的實踐慎思，道德化的思考和言論能更好地抵擋意志薄弱。道德話語也是做出人際承諾的良好工具（4.3節）。但是那個論證的核心部分是以下觀點：道德思維具有一種阻止算計和超越欲望的特殊實踐力量，這種力量使得道德思維通常比精明的務實慎思更能抵擋誘惑。然而道德自然主義的支持者不能采用那個論證，因為根據假設，他相信道德慎思就是關于欲求何物和如何滿足的。“這就不利于4.2節的論證了。”道德自然主義者也許會回應道[[111]](#_111_47)。但是，問題并不在于我提出的那個特定的論證，因為每當我們覺得“我們為什么需要一套單獨的道德話語？”有個好答案時，支持不含影響力的道德自然主義的人們都不能采用這個答案。唯一的結論就是這個道德自然主義者不得不回答道：“我們不需要。”作為道德的捍衛者（同樣也是道德實在論的辯護者），這個道德自然主義者不得不承認道德話語和道德思維沒有什么特定的功能，并且對我們的社會決策是多余的，這在他看來一定是極為難受的[[112]](#_112_47)。

以上反對道德自然主義的種種考慮，分開來看大概都不是致命的，但放在一起，它們至少還是揭示了這個理論的尷尬后果。首先，不含影響力的道德自然主義似乎削弱了我們在道德上批評犯錯人的能力；其次，它還可能鼓勵某些人犯錯；再次，它使得道德語言和道德思維完全成了多余的。這樣的價值系統顯然（再用一下之前的短語）太過虛弱，因而不會被誤認作是道德。我在前面論證了，道德思維有著特定的（既是演化上的也是當代的）功能，而這一功能是那些只涉及我們想要什么和需要什么的慎思無法實現的。我們實踐生活的道德化，通過準許懲罰，證成愛好和反感，以及把單獨的個人聯系在共同的決策框架里，從而幫助了我們的長遠利益，也促進了更有效的集體協商。我認為，正是因為道德規范所謂的權威性和不可避免性使其能夠實現這些功能。因此，缺乏實踐影響力的價值系統無法如此有效地扮演我們賦予道德的社會角色；所以我們不能像使用道德那樣使用這個系統，而這也就表明，實踐影響力可以被看作是一個對道德而言至關重要的方面（即便我們只是在粗略地說）。

在6.3節，我評估了道德自然主義滿足實踐影響力這一所謂必要條件的可能性，并且下結論說希望渺茫。在本節中，我探討了一種根本不想滿足實踐影響力的道德自然主義的前景，得到的結論也是一樣的悲觀。我最后的結論就是，實踐影響力確實是任何道德理論的必要條件，而且沒有什么形式的道德自然主義能滿足這一條件。這對道德自然主義可就更不利了。

## 6.5　重溫哈曼的挑戰

很多哲學家是全局自然主義者。要是沒有選擇某種非認知主義，他們所能認同的元倫理學立場就只能是某道德自然主義或者道德虛無主義。因此，對他們來說，如果以上的論證成功地否決了道德自然主義（并且非認知主義不在考慮之列），那這些論證也就成功地證明了道德事實根本不存在。需要注意，在奠定道德虛無主義的過程中，哈曼的挑戰和人類道德能力的譜系學都沒有扮演什么角色，可能頂多只是解釋了我們為什么一直以來都被系統地誤導以至于相信道德。

但是也有許多哲學家不是全局自然主義者。道德非自然主義者認為，道德屬性是存在的，但在某種意義上獨立于自然科學所描述的世界：道德屬性不等同于，不能還原成，也不依附于任何自然屬性。道德超自然主義者認為，道德屬性是存在的，但其存在依賴于某些超自然的現象（最顯著的例子就是上帝意志或命令）。即便這些立場在當代哲學家中不流行，但是我可以肯定，它們最為接近普通人的想法。因此哈曼的挑戰還有任務要完成。考慮圖6.3。回想一下，他的挑戰是假說A有希望解釋我們所有的道德判斷，因此除非道德事實以某種方式隱藏在假說A之內，我們就沒有必要預設任何道德事實的存在（即沒有理由假設我們有任何道德判斷是真的）。道德事實能這樣隱藏的唯一可能情況就是道德自然主義是真的（參見6.2節）。前面兩節內容給這種可能性打上了個大問號。于是我們就得面對圖6.3，而這次奧卡姆的剃刀真的可以起作用了，因為非自然主義和超自然主義確實給世界預設了額外的本體論，而非道德性的譜系學（假說A）表明這種本體論在解釋上是多余的。

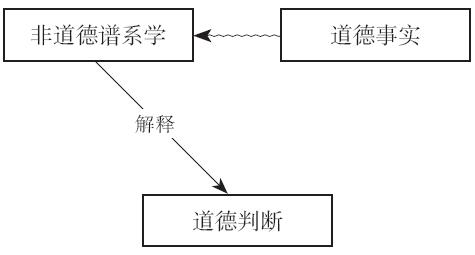


圖6.3

我在此細致地闡明不同立場之間的論辯關系，是因為我覺得這經常被誤解。前面提到了，哈曼的挑戰常被視作道德自然主義面臨的問題。這個挑戰確實要求道德主義者清晰地闡述自己的觀點，揭示自然領域如何同道德領域聯系起來，然而一旦他們的觀點清晰闡述了，挑戰也就消失了（事實上，很難想象自然主義者要是不能回應挑戰還怎么能把理論表達地有說服力）。要是自然主義者能夠有理有據，非自然主義和超自然主義就變得更不合理了。但要是自然主義者做不到有理有據，哈曼的挑戰似乎也會讓非自然主義和超自然主義顯得無用。換言之：一旦我們有了一套完整的道德判斷的非道德譜系學，如果道德自然主義成功了，非自然主義和超自然主義就沒救了，而如果道德自然主義者失敗了，非自然主義和超自然主義也一樣沒救。所以，在這場爭論角逐中，非自然主義和超自然主義遭受的損失最為慘重，而道德自然主義者只是因為額外的，跟哈曼的挑戰毫無關系的論證（即我在前兩節提出的那種論證）才敗下陣來。

這個論證的結果并不是道德虛無主義（道德虛無主義大意是說我們所有的道德判斷都是錯的）。指出我們沒有理由相信道德事實，并不意味著我們有理由不相信道德事實[[113]](#_113_47)。我沒有理由相信我的頭發數量是奇數（即便排除了模糊性的問題），但我要是因此下結論說有理由不信數量是奇數就很愚蠢了，因為那也就構成了相信數量是偶數的理由，而顯然，就像沒有理由相信是奇數一樣，我也沒有理由相信是偶數。雖然我沒有根據認為我頭發的數量就是奇數，但它仍然可能是奇數（其實是奇數的概率還不小）。讓我們回想一下信念藥片的虛幻故事。你發現了你那“拿破侖輸掉了滑鐵盧之戰”的信念只是某個跟他實際輸贏無關的過程的產物，于是你就不得不認為你既沒有理由支持也沒有理由反對這個信念。當時的結論是，你應該服下解藥，從而能帶著開放的態度去尋找其他更可靠的理由來相信或不信那個命題。（或者你也可以毫不在意，高高興興地在這個問題上保持不知情的狀態）。我從本章的論證中得出的結論也與此類似。對于我們的道德判斷如何起源，我們有了一個在經驗上得到證實的理論（我們是這樣假定的）。這個理論沒有直說也沒有蘊涵那些判斷是真的，也沒有把它們為真作為背后的假設，而且重要的是，它們之為真也沒有借助任何形式的道德自然主義而秘密藏在這個理論里面。這就等于發現了我們的道德信念是一個與這種信念的真假完全無關的過程的產物，因此我們就不得不承認，我們既沒有理由堅持也沒有理由反對這種信念。它們可能是真的，但是我們沒有理由這樣認為。因此，在最初的階段，我們應該培養開放的態度，去尋找其他更可靠的理由來相信或不信道德命題。

## 6.6　可靠主義者、保守主義者、融貫主義者和基礎主義者

雖然剛才得出的結論看起來很極端，但是有許多哲學家不會覺得擔心，因為他們認為我們相信道德命題是有合理基礎的。道德自然主義、道德非自然主義和道德超自然主義都可以說是形而上學的道德理論，因為他們的意圖是揭示那些使得我們道德信念為真的事實的本體論地位。與此相反，還有一些人嘗試純粹在認識論的基礎上維護道德信念。這些人的嘗試假定了我們可以證成道德信念，無論使其為真的事實的本體論地位如何。在這最后一節，我將證明道德的演化譜系學對于這些認識論上的努力確實提出了極大的挑戰。我將依次討論可靠主義、保守主義、融貫主義和基礎主義，其中主要篇幅用于可靠主義。

也許有人會論證說，給一個信念（或一個形成信念的機制）提供一套演化譜系學，不但沒有削弱這個信念，其實反而表明了這個信念是證成的。在一個名為“外在主義”（“externalism”，這不能跟4.2節提到的“動機外在主義”混為一談）的認識論傳統看來，一個人的信念可以由她主觀上意識不到的因素得到證成。過程可靠主義者（Process Reliabilists，外在主義最主要的支持者）就認為，一個信念是證成的，當且僅當它是一個讓信念和真理有可靠聯系的過程的產物。如果自然選擇是這樣的過程，那么任何先天的信念都是證成的。彼得·克勞瑟斯（Peter Carruthers）[[114]](#_114_47)大致就是這樣論證的：

……有些動物的行為大都產生于信念和欲望的交互作用，對于任何這樣的動物（比如我們）的生存而言，真信念都價值巨大。這是因為如果一個動物的計劃是基于真的信念，該計劃一般就會成功……長遠來看，完全錯誤的信念沒有促進存活的價值，而在演化選擇中長遠才是重要的。無可否認的是，如果生物的行為是基于真的信念，或至少近似為真的信念，那么長遠來看，它們一般都能夠存活。因此，如果某些先天的信念是從自然選擇中誕生的，那么我們就應該認為它們至少近似是真的……我們可以想象在某些情況下一個錯誤的信念有助于存活。例如，相信某種植物有魔力（實際是包含一種強效藥）的先天信念，對生活在那種植物生長繁茂地區的人們可能是有用的。但是一個信念要想被演化選擇，就需要在廣泛的環境里，同時也在相對人類歷史而言極長的時間里，都能發揮作用。一旦記住這一點，那么我們就可以知道上面想象的這種情況是不大可能發生的。因此我斷定，如果演化過程產生了任何的先天信念，那些信念都很有可能構成了知識（1992：111—113）。

然而，這個論證太過草率了。克勞瑟斯犯的錯誤體現了過程可靠主義的一個更深層的困難，即所謂的“籠統性問題”（Feldman 1986；Conee and Feldman 1998）。我們考慮一個具體的信念，并提問說產生該信念的那種過程是否可靠，但問題是這里說的究竟是何種過程一點都不清楚。這就好比指著一只狗，問它是哪種東西。正確的答案可以從非常籠統的（生物、脊椎動物等）到非常具體的（狗、臘腸犬等等），更不用提非生物性的范疇了（吵鬧的東西、被指著的東西，等等）。類似地，一個特定的信念既是籠統過程的產物也是具體過程的產物。設想一個具體的信念形成過程，比如艾米透過薄霧眺望，進而相信田野那邊有一頭牛。如果我們思考這個過程屬于哪一種籠統的類型，可能就會想到知覺（很多人覺得知覺是可靠的過程）。可是，這個過程也同樣屬于以下類型：視覺，清晨有霧時的視覺，艾米的知覺，以及艾米對于牲畜的知覺。這些信念形成的過程可能會不可靠得多（比如因為艾米是近視，或識別動物的能力非常差勁）。在最極端的情況下，相關的過程可能具體細化到只有一個個例：即我們當下談論的這一個。在此情況下，這種過程會被視作是可靠的，當且僅當這一個信念是真的，那么這就蘊涵著一個不可接受的結論：一個信念是證成的，當且僅當它是真的。

籠統性問題是否推翻了可靠主義，這還在爭論之中，我在此不會做出裁決（對該問題的有力支持，參見Conee and Feldman 1998）。顯然，要是這個問題無法被克服，我們就不需要再對這種思路做進一步的探討了。如果可靠主義者能夠解答籠統性問題，那就要提出一種選擇恰當過程類型的標準，既不太過籠統，也不太過具體。這就把我們帶回了克勞瑟斯對自然選擇是可靠的信念形成過程的簡潔論證。我認為他沒有找到適當的籠統程度。

為了討論克勞瑟斯的論證，可以把我的觀點解釋成：道德信念是先天的（實際上我是不會這么說的，除非加上很多限定）。我們可以很有把握地假定，擁有信念在很大程度上是因為信念的形成可以靈活地反映環境的變化。如果有些信念是先天的，它們肯定是特例而不是常規現象。因此如果有先天信念的話，調控它們的心理機制肯定不同于控制信念形成的常規機制。另外，不大可能只有一種生成先天信念的機制；更為合理的解釋是人類的心靈有一系列離散的心理機制，每一種分別都是被設計用來應對祖先環境里特定的威脅或機遇，而其中有些機制的作用是給個體提供固定的信念。比如有人論證過，人類對于語法，對于他人的心靈如何運作，以及對于義務和禁止，都有先天的信念。可靠主義者力圖從形成具體信念的過程中找出適當的籠統性程度，他們不想把這些過程之間的差異掩蓋了。產生關于語法的先天信念的過程，當然在一定程度上會同產生關于他人心靈的先天信念的過程有所重疊，但既然這些控制不同信念的心理機制是不同選擇壓力的產物，從而也是有不同工作程序的不同機制，那么相應的信念形成過程也是不同的。把這兩種信念看成是源于同一種過程（“自然選擇”），顯然就犯了過度籠統的錯誤。

以上論證迫使我們在考慮可能導致先天信念的過程時要更加仔細，逐個考察。在這里自然選擇的過程是無關的；自然選擇是個形成諸多機制的籠統過程，而那些機制的各種作用才是可靠主義者應該關注的具體過程。所以克勞瑟斯的假定，即所有的先天信念都是一種過程的產物，必須要否決。然而，他考慮的也不乏道理。也許把他的論證逐個用在具體的情況上，還是可以支持以下結論：相應的先天信念是可靠過程的產物。

克勞瑟斯論證的基本想法是，從長遠來看，錯誤的信念最終還是會令個體進行降低繁殖適應度的行為。如果我們考慮的是對于具體的環境因素以及各因素之間關系的信念，這個想法的確是非常有道理的。但是克勞瑟斯又給我們提供了一個可能的反例：相信某種有醫用價值的植物有魔力。因為魔力是縹緲難測的東西，有可能某些人對其抱有錯誤信念，但從來沒遇到實踐上的麻煩。同樣，現在相信鬼魂和祖先在天之靈的人們通常也不會在日常生活里遇到什么麻煩（這不是說這種信念就不愚蠢了）。相信世界起源的神話傳說，相信死后會如何如何，相信什么超自然的或神圣的存在俯視蕓蕓眾生——這些信念可能都對人類的決策沒有負面影響，甚至要是恰當地跟原本就有益（促進繁殖適應度）的目標結合起來，其實還可能有正面影響。克勞瑟斯考慮的情況就是如此：在某些環境里，出于醫學上的原因，吃下某種植物原本是有益的，所以再加上些“捉摸不定的涂層”（即魔力）可能不會給個體帶來麻煩，反而有可能促進繁殖適應度（比如要是魔力能加強吃下這種植物的動機的話）。他的論證指出，這種情況并不尋常，因此他還是可以說自然選擇造成的先天信念是可靠過程的產物，因為可靠的過程只需要在整體上產生真信念。但我之前的反駁指出，把所有東西都等量齊觀是不可接受的。如果某些植物有魔力的信念確實是先天信念，而這個信念是一個專門的、有特定演化歷史的心理機制的產物，那么我們需要關注的（如果我們是過程可靠主義者的話）相關信念形成機制就是這個特定的機制。因為對于這里的信念形成機制如何運作，我們可以得出一套得到經驗證實的解釋，該解釋甚至不需要預設那些信念是真的，所以，我們就可以得出結論說，任何這樣的先天信念都是不可靠過程的產物。而至于這是個不尋常的情況，以及許多其他自然選擇的先天信念形成機制的運作是可靠的過程，這些都是與我們的論述不相干的。

我們現在可以把這里的經驗教訓運用到先天道德信念的情況上。看上去比較明顯的是，在一些重要的方面，這種信念就類似對植物魔力的虛幻信念。我們已經看到了，前幾章簡述的演化假說完全沒有預設道德信念現在是或者曾經是真的。第四章展開的論證最后結論是：某些助人行為提高了生殖適應度，而把這些行為“道德化”增強了做出這些行為的動機。可以預期，一個相信每個角落都藏著捕食者的人，最終難免會遇上麻煩，但與此不同，把信念“道德化”的人似乎不大會遇上什么麻煩。道德上的事務足夠“捉摸不定”（應該加上：還足夠靈活），因此道德信念可以輕而易舉地通過日常生活的檢驗，即便是在很長的時間段和各種不同的環境里。假設一個人正確地相信（在環境C中）合作是符合她的長遠利益的。假設另一個人相信（在C中）合作是有義務要做的，相信無論她的欲望是什么都要這樣做，相信如果她不合作就得受懲罰（在她看來這些信念都是真的）。在什么情況下，后者那些道德化了的信念會導致她做出前者所能避免的、降低繁殖適應度的行為？在我看來，這樣的情況是不存在的。（需要記住，將社會慎思道德化并不意味著贊同無條件的或過度普遍化的合作；參見4.2節最后一段。）因此我斷定，即便從過程可靠主義者自己的角度看，一種有關先天道德信念如何演化出來的合理解釋也會自然地引向以下結論：這些信念在認識上是沒有證成的。

可靠主義認為，一個人即便沒有任何正面證據的支持，也可以正當地持有一個信念。另外一個贊同這種觀點的是認識保守主義。這種立場認為，一個人即便沒有任何正面證據的支持，也可以正當地維持一個信念，而這只不過是因為已有的信念都被假定是理性的（它們在“被證明有罪之前都是清白的”）。然而，任何合理的保守主義原則都不能無視之前提出的論證。保守主義的任意一種說得過去的版本，都要能容納以下常識觀點：一個信念要是符合某些性質就會變得可疑。瓦爾特·辛諾—阿姆斯特朗很好地概括了這些認識上的規矩：

……設想一下，我們各自把一組數字加起來，我算出了一個總和，但你求得的總和卻跟我不同。如果你是我的孩子，我在幫助你做功課，那也許相信我計算正確的就是理所應當的。但要是你的算術跟我一樣好，那么當我們得到不同的答案時，我們就需要再檢查一遍，看看是誰算錯了，然后才能有理據地相信自己的答案是對的。（2005）

出于這些考慮，辛諾—阿姆斯特朗提出了以下原則：“當人們意見不合并且沒有不受影響的理由傾向某一方時，需要有進一步的確證才能證成地持有信念。”其他的情形也引出了更多的原則，以及最終的結論：

當一個信念是片面的、有爭議的、情緒化的、被假象蒙蔽的、能夠以可疑的起源（即不可靠的或有爭議的起源）解釋的，那么……證成的信念就需要進一步的確證……這些原則，或者與其類似的原則，在大多數人看來都是有道理的，也是為我們共同的認識實踐所假定的。（同上）

辛諾—阿姆斯特朗列出的這五種性質，正是那些能夠削弱信念的任何表面證成的性質，它們的存在使得信念需要得到進一步的證成。這一點并不是在反駁認識保守主義，只是在澄清一個信念被預設的證成在什么情況下會被推翻。一個在得到經驗證實的理論，解釋了我們道德信念的起源，但是完全沒有蘊涵或預設這些信念是真的，這種理論的存在就是一種對預設證成的否決因素。這種情況屬于辛諾—阿姆斯特朗列出的“一個信念能夠以可疑的起源解釋”。（順便提一下，辛諾—阿姆斯特朗論文的目的是證明道德信念也體現出了推翻證成的其他性質。）

也許有人希望道德信念能從保守主義的認識論原則得到證成，但要是從中得不到的話，這些信念就需要詳細的辯護。許多人通常求助于認識融貫主義[[115]](#_115_47)。融貫主義認為，對于“納粹是邪惡的”的信念，只要能合適地容入一個其他信念構成的融貫框架中，就可以得到證成。（對于融貫主義的一個標準反駁就是：“納粹是勇敢的”，這個信念也可以適當地容入一個內容不同的但一樣融貫的信念框架；但是我們現在不用考慮這類問題。）然而，一旦我們意識到我們想借以構建一個融貫集合的信念不只包括對“納粹是邪惡的”信念，還包括（跟經驗研究關系密切的）關于這個信念的譜系的信念，事情就變得復雜了。假設這后一種信念認為“人類那些關于邪惡的信念，是來自我們祖先的社會生活環境，而這一譜系學完全沒有預設這種信念是真的。”這就足以削弱前一種信念原本享有的證成地位，因為這關于譜系的信念立刻促生了另一個融貫集合的可能性：在這個集合里，前一種信念沒有得到支持，而是被解釋消除掉了。另外，融貫主義者的信念集合大概也會包含一些在常識上不言自明的認識論真理，比如辛諾—阿姆斯特朗表述的那些，其中有的類似于“如果一個信念是由完全沒有預設其為真的過程產生的，那么人們就需要獨立的確證才能夠正當地持有這個信念。”要以最融貫的方式把這些信念歸到一起，就要斷定原來的道德信念需要明確的確證；因此，即便是從融貫主義者自己的角度看，原來的道德信念也是可疑的。

要是我們接受了本章早些時候提出的一些論證，那么從融貫主義的觀點看，道德信念的境況就更加糟糕了。如果我們確實都認為道德框架具有“不可避免的實踐權威”，并且只有把這種權威看作是關于理由的性質才能說得通，但是，對于理由是什么這一問題，我們最好的和最合理的自然主義理論完全解釋不了這種性質，而且同時我們又決定要做全局的自然主義者，那么，要達到融貫，我們在自己的信念系統里必須要有所取舍。究竟是在哪里犯了錯誤，我們可以有諸多選擇，而且也許都是同樣融貫的（我認為，這本身就構成了融貫主義面臨的一個一般性問題）。但是道德譜系的存在就給舍棄道德信念提供了理據，因為這樣一來，我們不光承認了錯誤，還能解釋那種錯誤的緣由（這種解釋還跟科學信念關系密切）。

融貫主義者的傳統論敵是基礎主義者，后者認為某些信念在認識論上是特殊的，不是由推論得到證成的。我想，這么說應該不失公允：經典的基礎主義最能表達大多數普通人對道德認識論的模糊想法，尤其是在最極端和最發自內心的情況下。納粹之邪惡難道不是明擺著的嗎？我們甚至一般都不會覺得有必要證成對“納粹是邪惡的”的信念，而要是遭到質問，我們大概會說類似“好吧，要是你就是不這么覺得，那就是你有問題了。”這樣的話。這種回答聽起來都不像是在說，那個信念在認識上應該是自我證成的，而更像是在指出，任何要求給出證成的人在道德上就很可疑[[116]](#_116_47)。如果后者是這種回答的意思，我們就要把注意力轉到另一個道德判斷上，（“對于‘納粹是邪惡的’這個判斷，我們不應該給它尋找明確的證成。”）于是現在的問題就成了：“這個判斷是怎么證成的？”

把經典的基礎主義運用到道德的領域，就產生了道德直覺主義：這種立場認為，人們持有某些道德信念在認識上是證成的，即便他們不一定能夠從別的信念推論出這些道德信念（粗略地說，就是主張某些道德真理是“明擺著的”）。但是我們在此可以再次利用從辛諾—阿姆斯特朗那里得來的經驗教訓，從而說明道德直覺主義錯在何處（事實上道德直覺主義是他的真正目標）。那些在常識看來會使得信念需要獨立確證的那些原則（信念是片面的、有爭議的、情緒化的、被假象蒙蔽的、能夠以可疑的起源解釋的）無疑都適用于“道德直覺”。下面的兩個例子旨在說明道德直覺如何被假象蒙蔽。辛諾—阿姆斯特朗引用卡內曼和特沃斯基（Kahneman&Tversky 1979）的研究：先給受實驗者介紹一個想象的情況，然后要求他們對600人里孰生孰死做出判斷。同樣的一個選項，是被描述為“200人會得救”還是描述“400人會死去”，會引起實驗者非常不同的道德直覺。另一個（辛諾—阿姆斯特朗沒有提及的）實驗表明，受實驗者在雙手從下往上頂桌子時，比起手放在桌面往下壓桌子時，更容易正面評價不熟悉的漢字符號（Cacioppo et al.1993）。對此的解釋假說（威廉·詹姆斯在1884年已經預見到了這種說法）認為，一個物體的價值得到何種評估，部分取決于運動的過程，尤其取決于肌肉是彎曲的（往回拉）還是伸展的（往外推）。有很好的經驗科學上的理由認為，我們發自內心的和貌似無可置疑的道德直覺經常都受到這些無關的因素影響。這就凸顯了一個問題：看起來是證成的信念，實際上可能并沒有證成。道德看起來是證成的，而且我們極不情愿承認它沒有證成，這些事實都沒有使道德真正成為證成的，其實這些事實本身就是譜系假說預料之中的現象。對哲學這門學科而言應覺尷尬的是，哲學史上充滿了這樣一些思想家，他們拿出自己偏狹的、偶然的、個人的意見，觀察到這些命題“是明擺著的”，隨即把它們上升到“自明真理”的地位。在他的自傳中，約翰·斯圖爾特·密爾提到了為直覺主義所鼓勵的教條主義的危險：“我相信，當前那些認為不依賴觀察和經驗就能由直覺或意識而認識到心靈以外之真理的觀念，給錯誤的學說和惡劣的體制在智識上提供了極大的支持。借助這一理論，每一種已經記不清由來的習慣信念和強烈情緒就可以免于用理性證成自身的義務，給自身構建起充分的擔保和證成。這種把根深蒂固的偏見神圣化的工具過去還未曾有過。”（1879：225—226）基徹爾（2005：176）的話也表達了大致相同的精神：“在倫理學里，就像在數學里一樣，訴諸直覺是一種絕望認識論。”要是沒有合理的理論說明道德直覺是如何自我證成的，這自然就會令人懷疑，其實真正需要解釋的現象是為什么這種信念在我們看起來毋庸置疑，并且不需要進一步的證成。這種解釋正是道德譜系學希望有所貢獻的。這種“解釋消除”的策略，要好過堅持那種本身就有問題的方案：力圖證明我們的道德信念在認識上是證成的。這是因為，在前者之中沒有殘留著神秘的現象或未解答的問題；它在解釋上是完整的。

最后的這些議論確實遠不夠全面，而且我也承認，一些復雜的問題和認識論理論還沒有得到討論。但我的主張是，如果經驗證據支持的理論完整地解釋了我們為什么會產生某個信念，而且該理論還不預設這個信念為真，那么在這種情況下，要是某個認識論理論還認為這個信念的證成地位并不受此影響，那這個理論就不值得擁護。任何在認識論上，令我們成為祖先繁殖能力的奴隸的理論，我們都應予以否決或修正。

# 結語　與適應的心靈共存

要是他真認為美德和罪惡沒有區別，那么，先生，不如等他離開房子以后咱們去數一數咱們的勺子吧。

——鮑斯威爾《約翰生傳》

倘若有一個不從事哲學研究的人，對他人行為舉止的了解大都是來自于約翰·麥基的行為，那么他做夢也不會想到用“道德懷疑主義”來解釋這些行為。

——喬治·考克威爾，約翰·麥基紀念致辭，大學教堂，牛津，1982年2月27日

1883年，托馬斯·赫胥黎在牛津大學做了羅曼尼斯講座。在維多利時期的英格蘭，因為他對演化論的激烈辯護，赫胥黎以“達爾文的斗牛犬”的稱號而廣為人知。《牛津雜志》提到了“華美的博士袍下講者令人難忘的身形，那身袍子極好地襯托出他精致的輪廓和敏銳智慧的頭腦上面長長的白發。”盡管演講經常被大廳外面石子路上的馬車嘎嘎吱吱的響聲淹沒，赫胥黎還是給聽眾留下了難忘的短句：“讓我們就此徹底明白，社會的倫理進步，并不在于模仿宇宙的進程（即自然選擇導致的演化），更不在于回避這一進程，而在于與這一進程做斗爭。”［（1893）1947：82］。一個世紀之后，喬治·C.威廉姆斯仍然贊同這一想法，給他的一篇文章取名為《自然母親是個邪惡的老婆娘》（“Mother Nature is a wicked old bitch”，1993；參見Dawkins 1981：3）。

我們不應該忽略這些態度不經意透漏出的反諷意味。如果我們的道德感受力本來就是生物自然選擇賦予的，那么要是沒有自然選擇，我們就無法思考“社會的倫理進步”。假設赫胥黎和威廉姆斯都同意，擁有進行道德判斷的能力（乃至使用這一能力來評價自然選擇的過程）是值得贊賞的，那么自然母親能有多邪惡呢？他們的論證是搬起石頭砸了自己的腳。衷心地跟自然選擇“做斗爭”就需要抵制道德感本身（在努力爭取我們加入他的事業時，赫胥黎希望激發的偏偏正是這種能力）。

自然選擇不應受到赫胥黎和威廉姆斯那樣的批評。自然選擇的過程讓我們能夠合群，能夠參與合作性的交換，能夠去愛，能產生共鳴和進行利他行為（讓我們有能力直接關心他人的福祉，而不考慮回報），還把我們設計成能夠用道德觀念來看待自己和他人的關系。為什么自然母親對我們如此慷慨？倒也不是出于什么太值得稱道的目的（所以我們也無須過分贊美她），無非只是因為為人友善可以幫助我們的祖先繁殖更多后代。如果覺得這些親社會的傾向只牽涉非認知的感覺、行為傾向、意愿、厭惡和偏好，而不涉及信念，這是幼稚的。但承認了信念受到自然選擇的影響，這也就引發了認識論上的擔憂。因為當繁殖的成功才是標準時，準確地表征現實世界的只有偶然的工具價值，所以我們就要承認，假如在某些領域里，錯誤的信念能帶來更多的后代，那自然選擇就會每次都往這個方向走。道德思維很可能就是這樣的領域。

很多所謂的“規范演化倫理學”都是從人的道德感是演化產物這一前提出發，然后試圖從中提煉出某種道德結論。人們經常很有信心地回應道，單憑描述性和歷史性的信息，不論有多少，都無法告訴一個人他應當做什么。但是我在上文論證了，演化倫理學也有更深刻也更令人擔憂的一面。要是這些描述性和歷史性的信息是關于他那種追問自己應當做什么的能力本身呢？假如這些信息揭示了他用以進行慎思的概念，其特性跟真理無關，而是來自10萬年前非洲大草原上的社會狀況，結果會如何呢？假如他意識到了，自己對質疑道德的那種抗拒（他那在哲學課堂上都想要為道德辯護的動機），本身也可以在同一個解釋框架里得到說明呢？這些問題的答案，這些純粹描述性的信息并不是告訴一個人應該做什么，而是這些信息對他的慎思有破壞作用。這里用“破壞”不是想表達貶義；如果你在某些問題上的想法，表面上好像是對于世界的準確表征，但其實并沒有理由認為世界確實如此，那么，按照認識的標準，動搖這種想法就是件好事。

我在本書論證了，有關道德譜系的描述性知識（再結合一些哲學思辨），應當削弱我們對于自己的道德判斷的信心。這一結論可以被視為一種道德懷疑主義。這里說的懷疑主義，在這個術語傳統的意義上，是指證據既不支持也不否定某一命題，因此至少在新的證據浮現之前，我們應該選擇（或者被要求）在相應問題上懸置判斷。這就是我在書里偶爾提到的，在相應問題上“培養不可知論”。在宗教領域，不可知論這一觀點內容是指我們無法知道上帝是否存在（這也確實是該術語的發明者托馬斯·赫胥黎的本意）。但如果成為有神論者就等于是相信上帝存在，而成為無神論者就等于是相信上帝不存在（注意這兩種立場都沒有提到知識），那么不可知論要與前兩者鼎足而三，我們就必須將其理解為既沒有相信上帝也沒有不信上帝的立場[[117]](#_117_47)。無論如何，這是我對“不可知論”這個詞的用法。另一種道德懷疑主義者則類似于無神論者，宣稱我們有好的理由相信道德判斷都不是真的。本書中的內容有可能包含了可以用來論證這一更極端的懷疑主義的內容，但我在此并沒有想要如此論證。元倫理學的標簽“錯誤論”（“error theory”，這個術語是由其擁護者約翰·麥基發明的）常常僅限于用在無神論式的道德懷疑主義者身上，但沒有什么阻止我們把它的用法拓寬到本書所青睞的懷疑主義立場：道德判斷在認識上都沒有得到證成（關于這一術語的拓展，參見Kalderon 2005）。因此，只要心里清楚這兩者的分別，我很樂意給本書論證的觀點貼上“道德錯誤論”的標簽[[118]](#_118_47)。這兩種錯誤論有一點是共同的：兩者都不相信道德命題（一個是因為道德命題不是真的，另一個則是因為我們沒有權利贊同道德命題）。

我估計，不熟悉元倫理學爭論的人們會對這一結論感到極為震驚。即便那些習慣了相關爭論的人，可能也會覺得這個結論令人震驚，但至少他們不會對有人為其辯護感到驚訝。對于這種反應，我有兩點想說。

第一，道德懷疑主義者很難被針對她立場的道德譴責挫敗。如果一個人想要否決任何形式的道德懷疑論，但他的動機基礎是道德上的考慮，那么被他審查的理論就可以回應道：那種考慮是錯的或者是沒有證成的。任何一種旨在說明普遍存在的錯誤的理論當然會是違反直覺的。因此，道德懷疑主義者本來就會預料到，他的觀點在很多人看來是明顯錯誤的和有害的，所以不能用這種看法作為反駁其觀點的考慮因素。如果僅僅是心懷震怒地對道德懷疑主義者的結論指指點點，其實只是回避了問題而已，對于反駁論證毫無幫助。

第二點，也是更為重要的是強調這種道德懷疑主義在實踐上蘊涵的后果是非常不明確的，以此來緩解一下那種震驚的反應。約翰生博士（見本結語的開頭引文）擔心邀請道德懷疑主義者來家里喝茶可能危及家里的餐具，我認為是不正確的。道德懷疑主義者認識到偷竊行為沒有或可能沒有任何特定的道德錯誤，但他的觀點完全不是在主張普通人毫無理由的避免偷竊（把這截然不同的兩件事混作一談是巨大的錯誤）。否則就等于是在承認，橫亙在我們和野蠻放肆偷勺子的生活之間的，也只有道德責任感而已——這種想法實在令人沮喪。對某一個特定的規范框架提出質疑并不意味著“怎樣都行”。作為參照，在更早的時代，有人擔心宗教無神論會導致社會秩序的崩潰（要是不相信有神明監視著人間事務，人們一旦確信能夠逃脫責罰，就沒有理由避免惡行）。雖然我敢說現在有些人還是這么想的，但我們大多數人都意識到這是毫無根據的焦慮。（無神論者占人數比例最高的國家依次是新西蘭、挪威、烏拉圭，它們都不以道德墮落著稱；美國可以自夸是世界上宗教活動參與度最高的國家之一，但犯罪統計數據也最為糟糕。）就算沒有憤怒的上帝在旁邊盯著，就算沒有神秘的道德責任禁止偷竊，就算她能夠逃脫懲罰，一個普通人有許多理由避免偷竊（簡單而有力的理由）。（其中一些理由的根據在于，人類被設計成為社會性的動物，而對于社會性的動物而言，真誠地參與社群能帶來重要且不可替代的滿足感。）這樣說并不是否認相信上帝或道德責任能增強動機（也許實際上這正是這些信念的核心社會功能），而只是想指出，我們一般都具有獨立于上帝或道德責任之類因素的理由，來進行“親社會”行為。即便宗教信念和道德信念確實促使我們的親社會動機和親社會理由保持一致，我們也沒有理由認為人類不能形成或強化某些其他的社會和心理機制（那些跟宗教或道德信念毫無相似之處的機制）來達到一樣的良好效果。

另外，從認識的角度禁止道德信念，這對于道德情感的地位有何影響，也是不明確的。如果這些情感必然包含著道德信念，那么處于認識上的理由它們也會被禁止。但我在第三章花費一番工夫，論證了就連那些充滿認知成分的情感也不需要信念——光是想法就夠了。而第六章的內容，沒有什么能構成反對某些特定想法的認識上的考慮。（相信你是世界上最好的網球選手幾乎肯定在認識上是沒有證成的[[119]](#_119_47)，但光是有這種想法，當作無傷大雅的幻想，也沒有什么問題。）讓我試著通過跟一種非道德情感進行比較來澄清這個問題。這種情感是接觸某些物體時感到的惡心。在兒童時期，對于傳染性污染（contagious contamination,“cooties”）的觀念的固著（f ixation）似乎是一種幾乎人人都有而且經常是自發的現象。這種現象在沒有得到相應文化的支持時會逐漸消失，但也可以在文化中表現得非常精致繁復，成為核心的社會原則。（參見Rozin and Fallon 1987；Rozin and Nemeroff 1990；Kalish 1999；Greene 2002：195—196；Hejmadi et al.2004。）這種情感似乎有先天的基礎，可能是自然選擇為了產生某些回避的動機而形成的，這些動機的產生不完美地對應于人類祖先時期面臨的一種真正的威脅：病原體傳染[[120]](#_120_45)。我們的問題是：理性的成年人應該怎么看待cooties？

“cooties”這個詞最初指的是體虱，在書面用法里也經常指細菌；但我在這里遵循的是北美地區的用法，泛指某些看不見也說不清的有害污染物（比如，“girl cooties”）。后一種意義上的cooties常常引起惡心，而病原體本身一般不會有這樣的作用。注意以下兩種感覺的區別：一種是即將從最近剛裝過冷凍細菌的壇子（但已徹底清潔過）里喝水，另一種則是即將從最近剛裝過福爾馬林液，用來展示畸形胎兒的壇子（但已徹底清潔過）里喝水。（我的意思不是說人們會更不愿意選擇后者，只是說這兩種引發的感情是非常不一樣的）。要是我們熟悉了這種惡心感是怎樣演化出現的，我們應該怎么看待這些情感反應？如果面臨著要穿上一件剛洗過的、曾經屬于某個殺人犯的毛衣（Rozin et al.1994），或者用一個嶄新且干凈無瑕的便盆喝蘋果汁（Rozin et al.1999b），對于這些情景感覺到的惡心有多么理性呢？理性而有知識的成年人當然不應該相信這些物體可以傳播什么看不見的污染物，然而他們可能會發現無法把這種幼稚的想法從腦海除去。光是便盆這個念頭本身就會讓你想到某人把它用于它原本的用途。即便你知道從來沒有人用過這個便盆，從一個真被弄臟的便盆里喝水，這個念頭實在令人厭惡，你都沒法把它拋之腦后，就好像一個人開車經過一起嚴重車禍現場時就是想要去窺探一番。也許這個想法和由它引發的原始抗拒感就足以讓你對于從干凈便盆里喝水這一情景感到惡心。但你不一定就因此而犯了任何認識或理性上的錯誤。但是情感可以是非理性的；我曾經認識一個在其他方面都算聰明的人，但她極其害怕氣球。然而，無論用來判定情感的非理性的框架是什么，它都不同于用來判斷信念的非理性的框架。也許有人想把不符合相應信念的情感稱為非理性的，比如也許有人覺得我朋友對氣球的恐懼癥（pallophobia[[121]](#_121_45)）是非理性的，因為她并不真的相信氣球是危險的。然而這肯定是不對的，因為讀者對于小說的情感反應也跟相應的信念不一致，但（就我所知）只有一個哲學家愿意稱他們是非理性的[[122]](#_122_45)。被歸入非理性的情感狀況一般都帶有強迫和自毀的性質。但從干凈便盆里喝水的情景而導致的惡心無須具有其中任何一種性質。

另外，對于理性的人而言，這一情感的反應本身還可以代表一個不從干凈便盆里喝蘋果汁的理由。要是他渴了，或者是別人給他十塊錢讓他喝，這個理由的分量可以被輕易蓋過。但是假如沒有支持他去喝的重要反對理由，那么他不想喝（無論這種不情愿是因為他無法把某些討厭的畫面拋之腦后，還是因為某種模模糊糊的怪念頭）這件事本身就構成了拒絕的有力理由。我想指出的是，即使一個人遵循了認識上的要求，從腦海中去除了所有關于那種“一旦有接觸，永不能祛除”的神奇關系的信念，但要是她繼續有這樣的念頭，或感覺到與此念頭相聯系的惡心，或允許這一感受在自己的實踐慎思中占有一定分量，這些都無須算作是認識上或理性上的失敗。

這些看法也適用于道德情感。即便某人遵循了認識上的要求，堅持道德的不可知論，但要是她的想法還是包含道德概念，或她還具有一整套的道德情感（雖然沒有信念），或者她允許這些情感在自己的實踐慎思中占有一定分量，這些都無須算作是認識上或理性上的錯誤。沒有人否認情感能夠影響動機（也許有的人還會說情感必然影響動機），而如果人可以具有某些情感但缺乏平常與其相聯系的信念，那么即便沒有那些信念，也同樣可以引起動機。更一般地說，我們需要記住，有充分的證據表明影響我們動機的因素可以是微妙的，驚人的，乃至多少有些莫名其妙的。圣奧古斯丁都已經注意到適當的肢體動作可以加強情感（比如俯身在地增強了虔誠感），但他隨即補充說他也不知道個中緣由［Agustine（c.426）1968：360］。現代心理學的實驗為他的觀點提供了支持。自發的微笑確實對情感有積極的影響（Ekman and Davidson 1993），而一個人看喜劇節目時感到的快樂，可以被他嘴叼著鉛筆的方式影響，因為鉛筆在跟牙交叉的情況下會激活微笑用到的肌肉（Strack et al.1998）。同樣是被要求把詞序打亂的句子排列好，面對的句子是關于老齡主題的那些受試人，比起排列關于其他主題的句子的受試人，在做完實驗離開的時候要走的更慢一些（Bargh 1992）。一個剛在公用電話的退幣口撿到一毛錢的人，幫路人撿起掉在人行道上的文件夾的概率要遠遠高于那些沒找到一毛錢的人（Isen and Levin 1972）。在城郊的路上幫助貧困的陌生人的可能性，會受到周圍環境有多少噪音的影響；比如，周圍有一臺汽油式割草機會嚴重降低提供幫助的概率（Mathews and Canon 1975）。我提及這些林林總總的研究，只是為了證明，人的動機比我們通常認為的要怪異得多，實踐選擇可以受到各種各樣的影響。也許讓道德想法和情感在人的心理秩序中扮演活躍的角色，即便沒有與之相聯系的道德信念，同樣也是引起動機的一種不明顯的方式。這些想法和情感或許會成為習慣，甚至融入個人性格之中[[123]](#_123_45)。我們沒有明顯的理由懷疑，這些想法和情感在人的生活中相當重要，甚至還能成為個人的或人際關系中的承諾。

我們還可以更進一步，不但說人們可以繼續讓道德想法和道德情感在他們生活中對動機施加影響，還說很多人應該這樣做。如果我在第四章的說法（在一般人的生活中，道德思維起到了正面的工具作用）大體是對的，那么從人的心理中徹底清除道德，要付出的代價是很大的。如果在形成偏好和制定決策的過程中，沒有信念的道德情感能夠起到積極的作用，從而減小這種代價（而且為了論證我們假定沒有更好的方式減小代價），那么人就應該讓這些情感起作用（這里是務實的“應該”）；我們甚至可以說，如果一個人不這樣做，那在實踐上就是不理性的。以上這些都可以成立，但同時她仍然可以是個道德懷疑論者。

如果我們本質上就無法遵從這些關于我們“應當做”什么或“不應當相信什么”的教誨，那它們就會毫無意義。但是，我在導言中說過了，有關先天性的說法并不意味著不可避免性。人類的道德感明顯極容易受到學習和環境的影響（而且可能這正是設計的結果），因而它似乎沒有什么不可避免的地方。就像兒童似乎自然而然地對某種污染物持有特定的信念一樣，道德信念要是沒有文化的支持可能也會逐漸消失。也許這看起來不大可能，而且沒有益處。然而類似的主張，比如所謂會因為觸碰下賤種姓的人而遭到污染，在某些人看來也是一樣的不大可能和有害，因為在他們深浸其中的文化里，這種純潔和污染的觀念已經成為個人、體制和國家的相當繁復的規范基礎。我們能不能通過自我訓練，不再認真地贊同道德信念，就像通過自我訓練不再相信某種污染物一樣？如果有了必要的文化支持，是很有可能的。但是即使答案最終是否定的，這也不意味著我們就是機器人，總是無法理解道德在認識論上的地位。休謨說“要不是自然的抵制力太強大，哲學就會讓我們全都變成皮浪主義者（Pyrrhonian）[[124]](#_124_45)”，這也許沒錯，但這本身并沒能駁倒懷疑論者的論證。

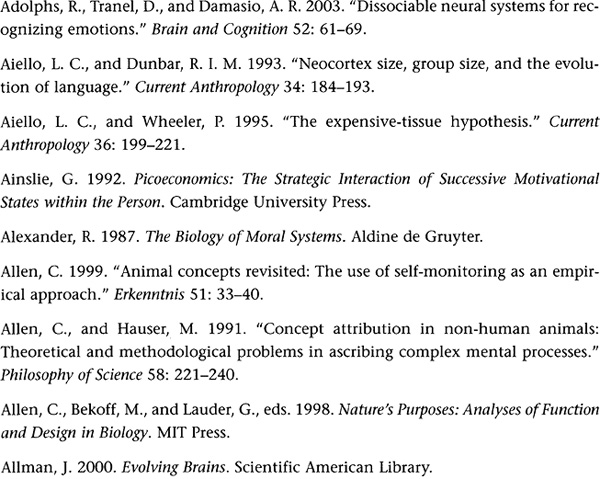
休謨在那本18世紀的道德心理學杰作中，解釋了人類道德的起源。他滿足于將其歸結于人類具有自然的同情心這一單純的事實：

我們的研究沒有必要更進一步，去追問我們為什么有仁慈之心，或者為什么能把別人視為同伴。知道這是我們能體驗到的人性原則就夠了。我們對于起源的考察必須得有個終點；每種科學都有一些普遍的原則，想再找出一些更為普遍的原則是沒有希望的事。［（1751）1983：43］

休謨的解釋沒有更進一步，因為他力不能及。但在他死后不到百年，達爾文就帶來了可以推進對人類道德的研究的工具，他給我們提供了一套理論，來處理“自然”（在休謨的著作里，“自然”經常是用以解釋其他事物的要素）為什么偏偏把我們塑造成這樣的問題。但是，雖然達爾文提供了一套值得尊重也可以驗證真假的道德譜系學，他對這套譜系學在元倫理學上的含義卻言之甚少（這或許也可以理解）。一方面，他完全了解，我們的道德本性是一個特定演化歷程的產物，這一事實就使得道德再無法稱得上客觀。在一次奇思異想中，他設想了假如我們是從類似蜜蜂的生物演化而來的，那我們會是怎樣的。他的結論是：“未婚的女性會像工蜂一樣，把殺死自己的兄弟視為一項神圣的職責，母親會努力殺死不育的女兒；而沒有誰會想要阻止這些行為。”［（1879）2004：122］但同時他又似乎很信賴自己那套維多利亞時代的道德觀念，輕松愉快地提到了“野蠻人的道德低下”，把奴隸制稱為“嚴重罪行”，自如地使用“高貴”、“邪惡”和“野蠻”這些詞匯。很不清楚的是，達爾文為什么覺得由大英帝國特定的文化進程所塑造出來的人的道德感，就可以產生真的或證成的道德觀念（即便正如他所言，所有大規模群居的人類最終都會具有這些觀念）。他的一條評論指出，那些涉及他人福利的道德規則“受到我們人類同胞和理性的認同。”（同上；147），但他并沒有展開闡釋，讓人空有期待。他的意思是不是說，以上設想的那些像蜜蜂一樣的生物，它們殺害兄弟姐妹的行為就是非理性的嗎[[125]](#_125_45)？

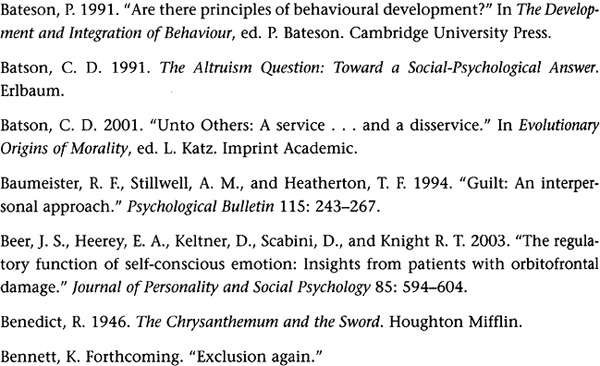
有些跟達爾文同時代的人意識到了他的譜系學給道德造成的威脅。弗朗西絲·柯布指責他的道德感演化觀是“馬基雅維利之后出現的最危險的學說”（1872：11）。在她看來，對于道德良心的達爾文式解釋“為了……給倫理學致命一擊，于是宣稱我們的道德感不僅來自不值得特別尊重的起源，也不符合任何外部的或持久的（更別說是普遍的或永恒的）客觀實在，還宣稱我們的道德感只不過是臨時的和暫定的，是（我們可以這么說）這個小小世界及其居民的地方偏見，要是讓在火星上那些更有見識的人看到了，他們只會報以嘲弄的微笑。”（同上：10—11）但是大部分這樣懷疑達爾文的道德演化觀有害處的人，他們的推理方向都反了：因為一些模糊的所謂的危險，他們就想要抵制和否定達爾文主義的理論。自不必說，這種態度完全應當反對。無論如何，至于現在討論的個例，我已經論證了，擔心我們生活方式會遭到嚴重破壞是沒有根據的。但即便這種擔心成立，唯一一種誠實而有尊嚴的應對方式，就是承認證據和我們最好的理論推斷所揭示的結論，并處理相關的實踐后果。［連柯布都承認，如果科學真能證實達爾文“危險的學說”，那就必須“坦然面對其后果”（同上；11）。］如果令人難受的真相就擺在那里，我們應該找到并像理智的成年人一樣面對這些真相，而不是避開虛心開明的研究，也不是編造些什么哲學理論，其唯一的優點就是有希望帶來令人寬慰的好消息：我們所有發自內心的信念都是真的。

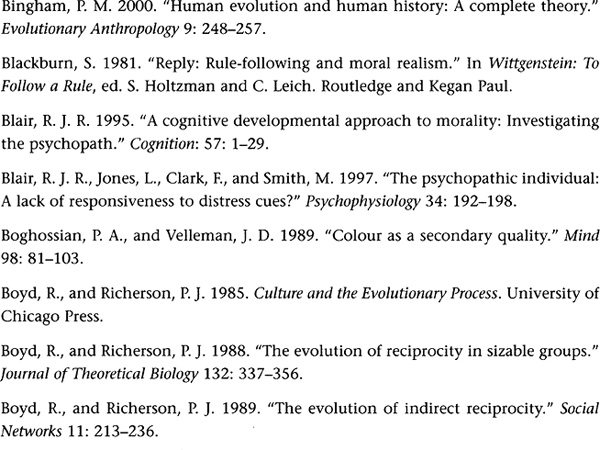
# 參考文獻

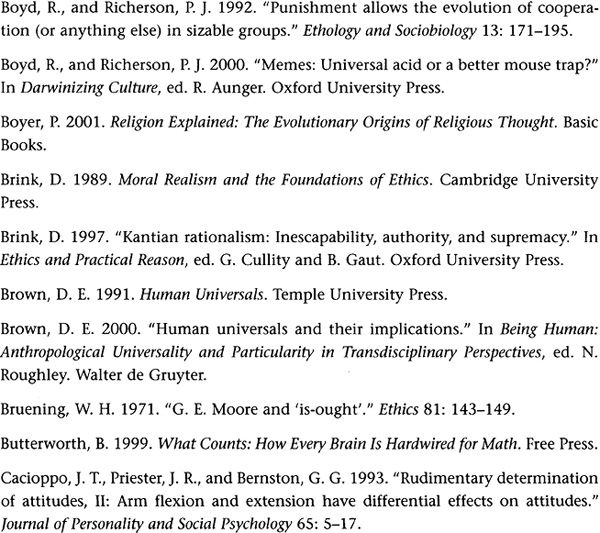


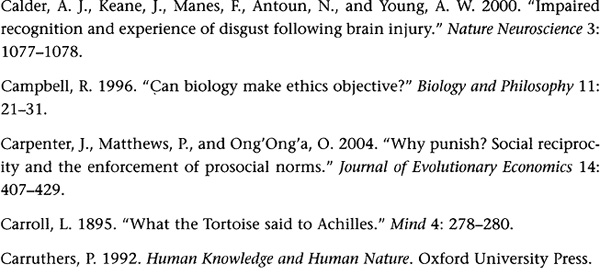


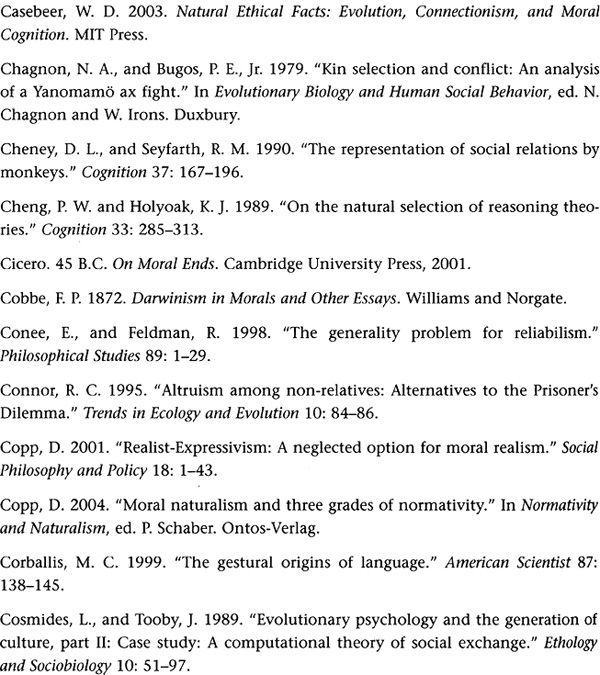


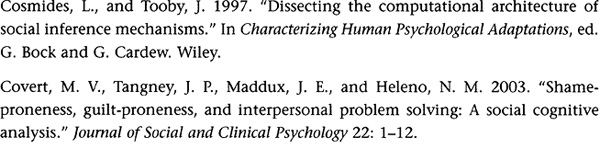


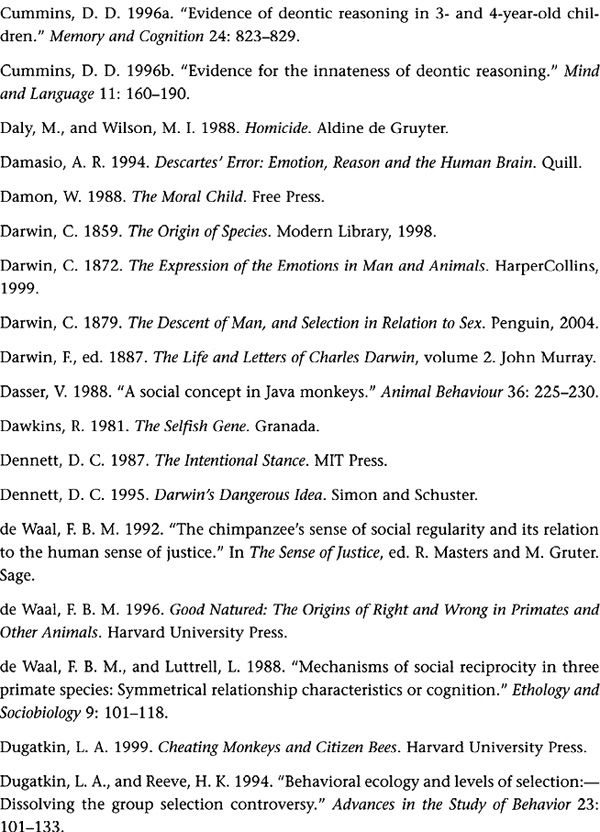


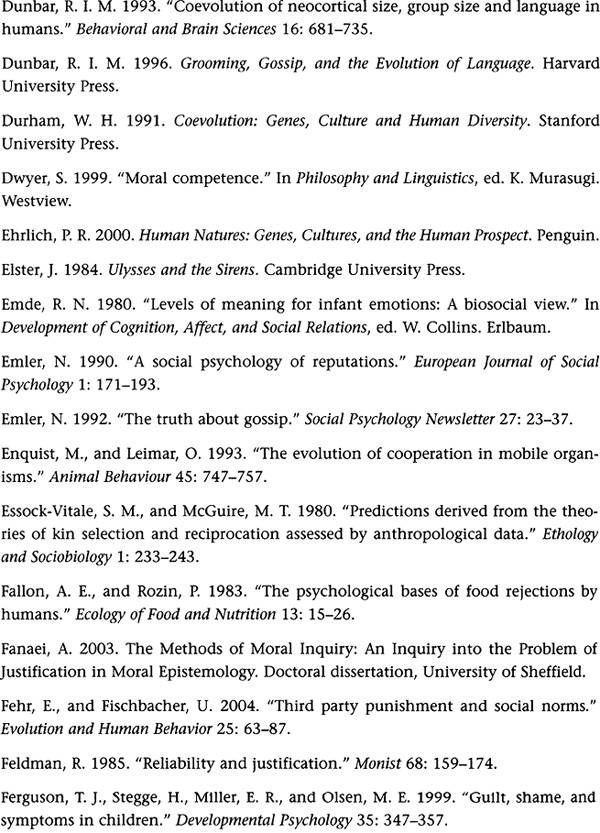


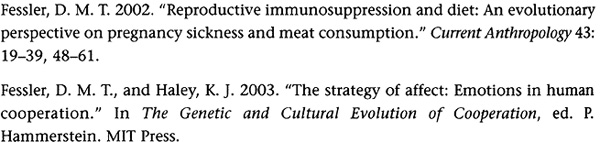


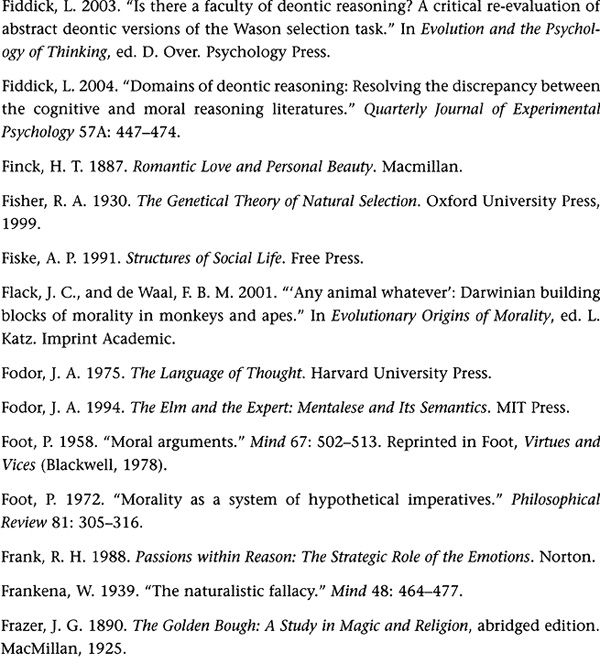


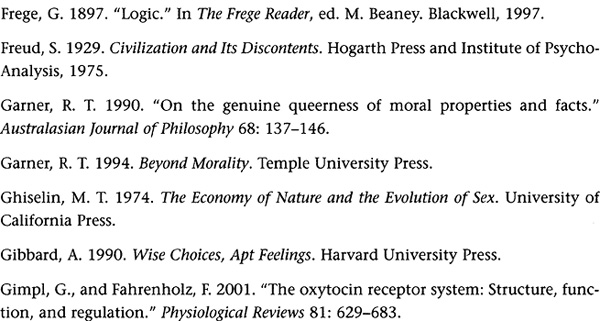


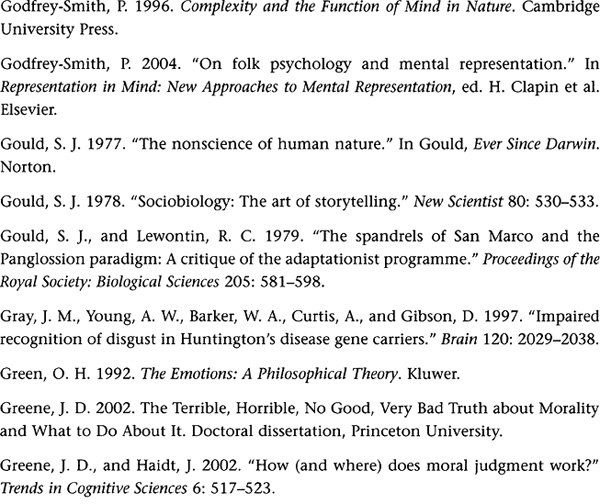


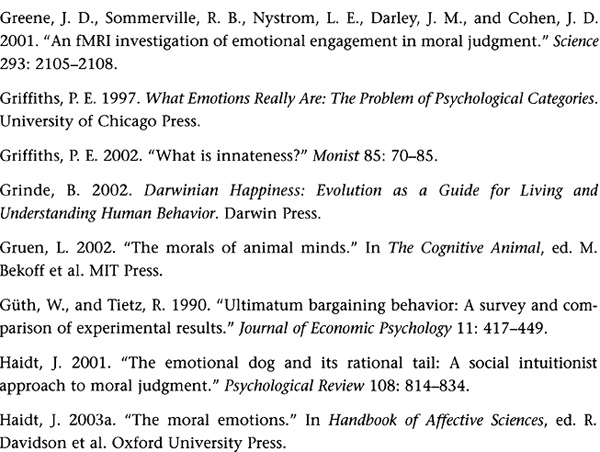




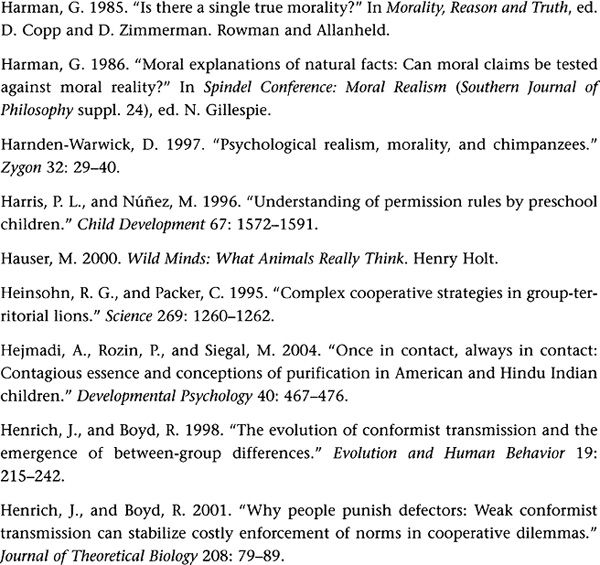


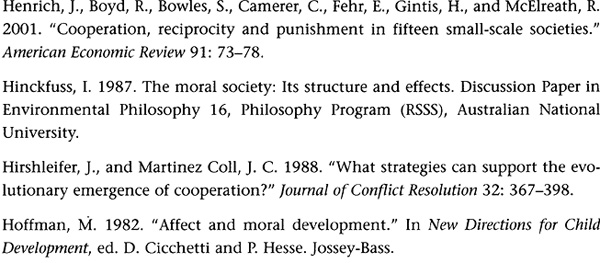


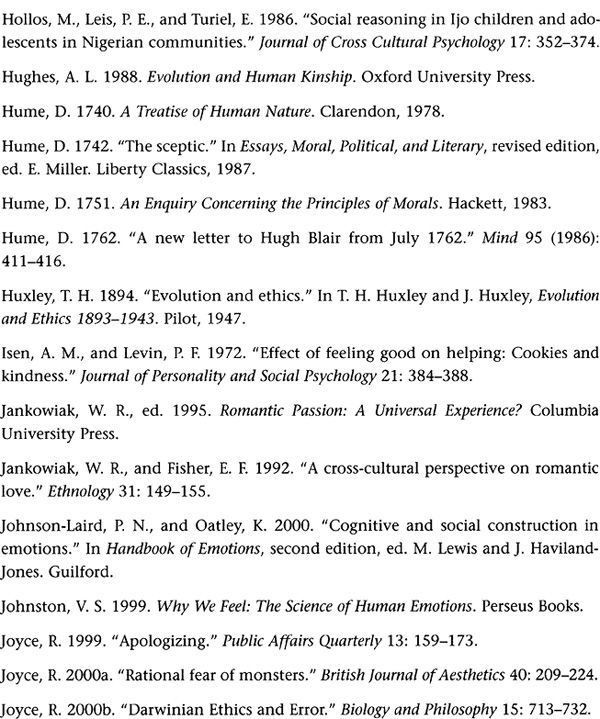


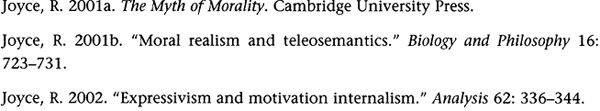


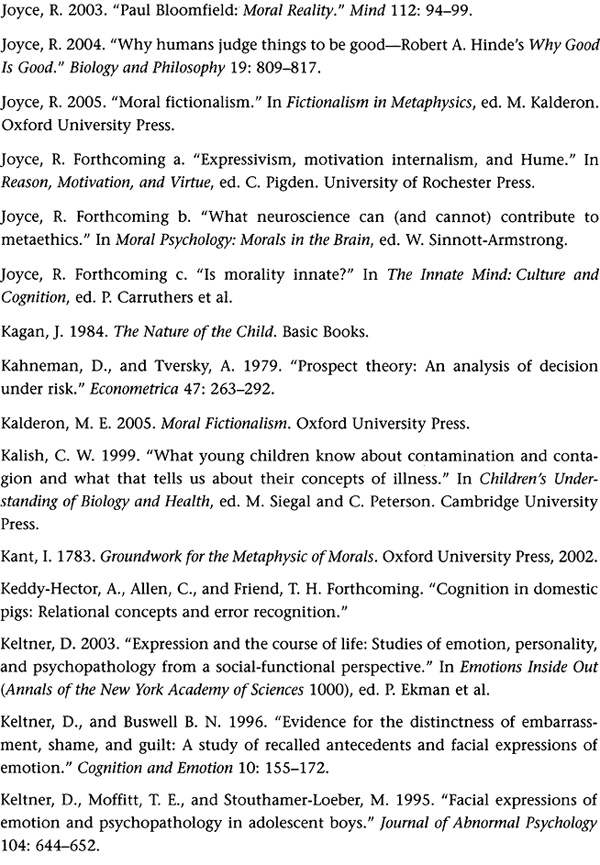


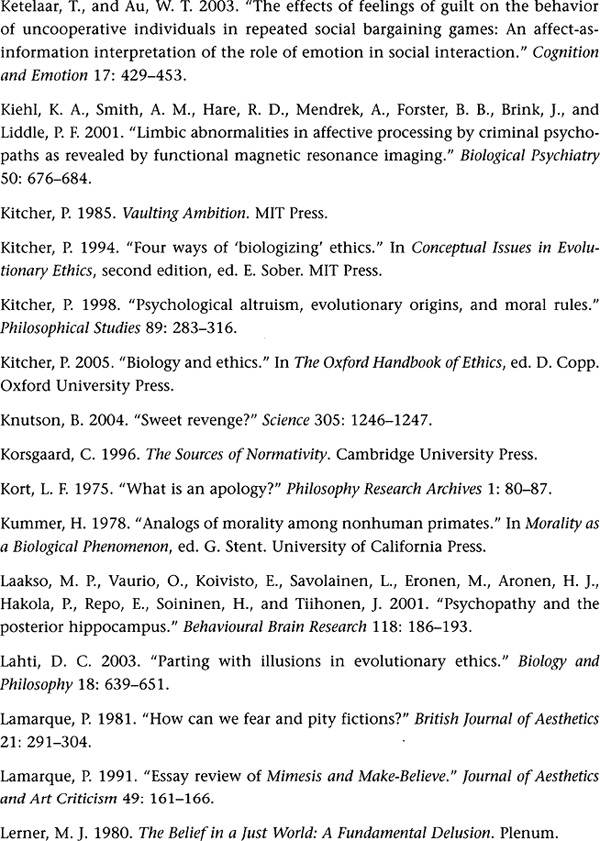


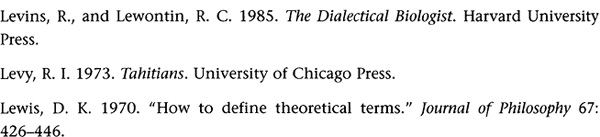


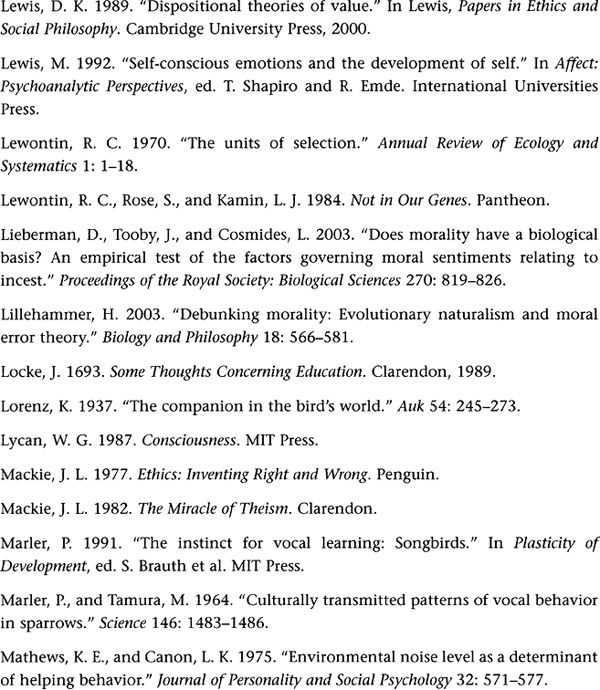


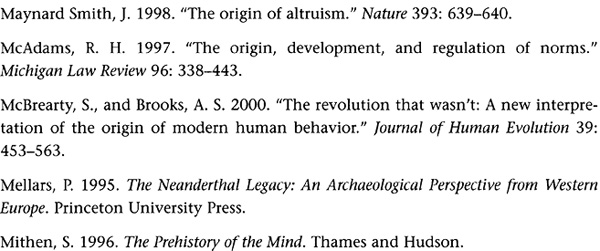




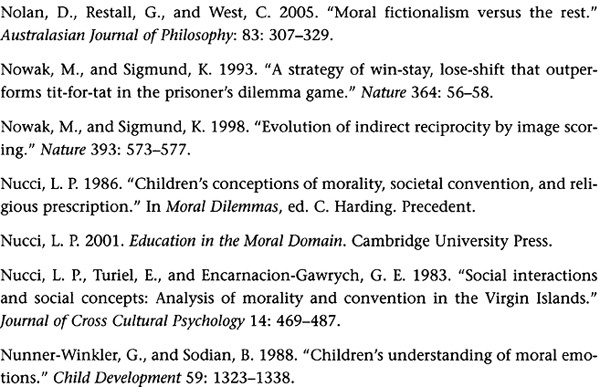


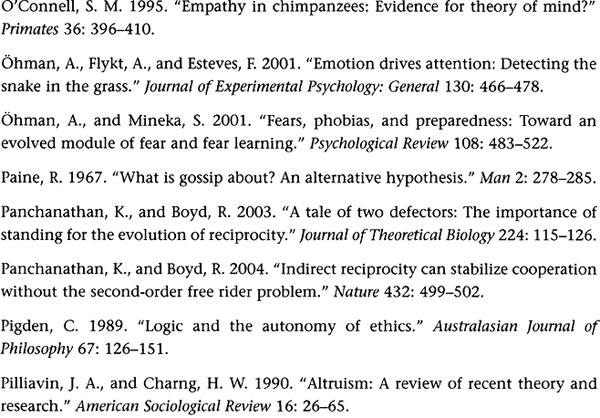


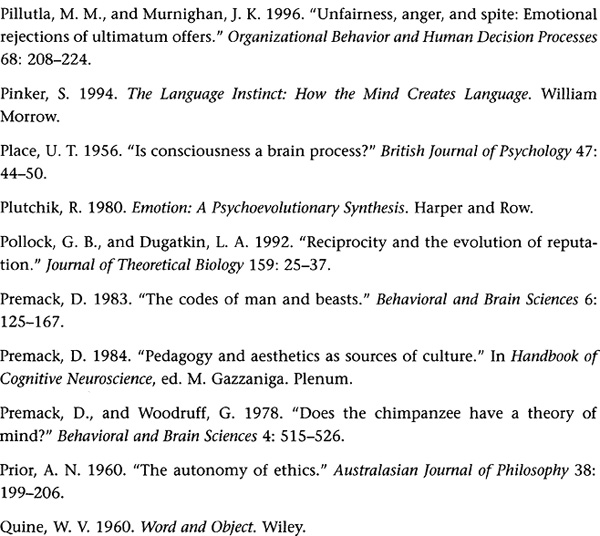


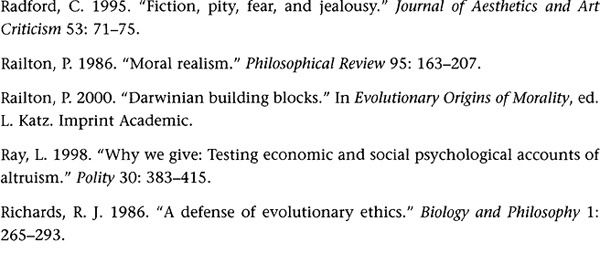


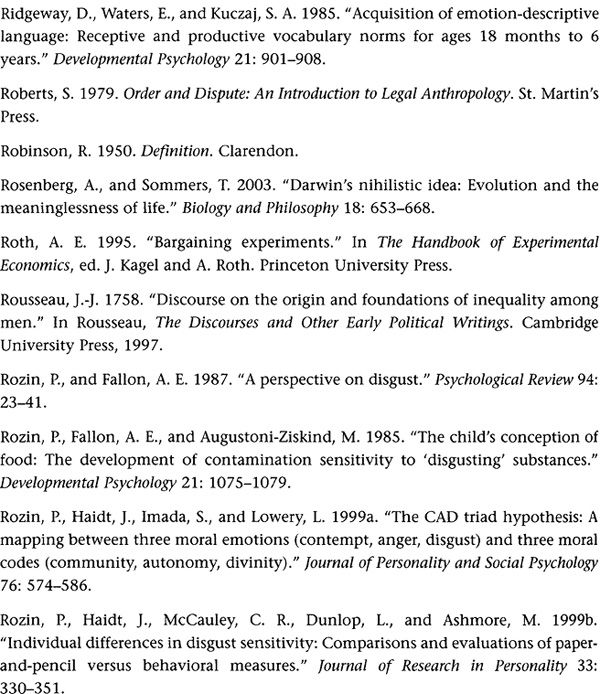


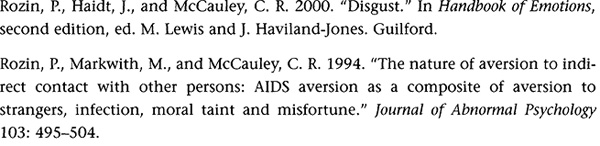


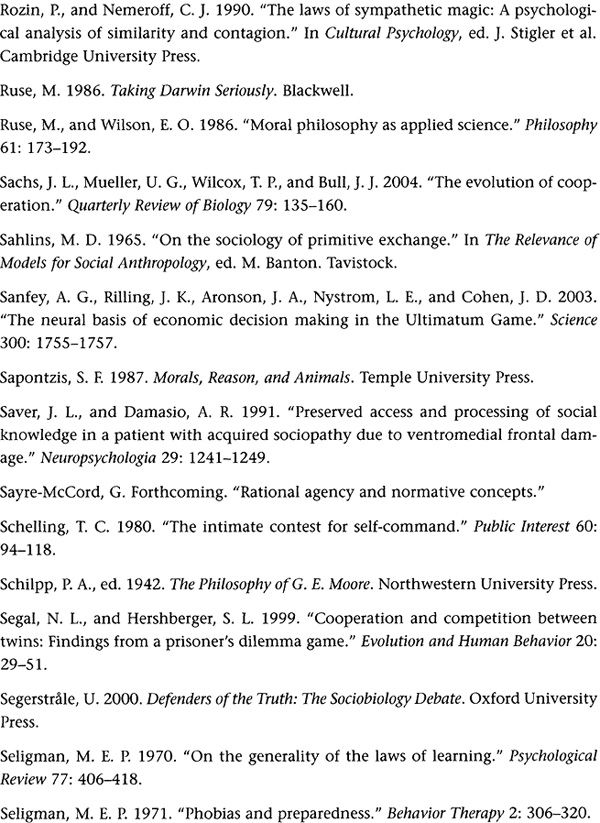


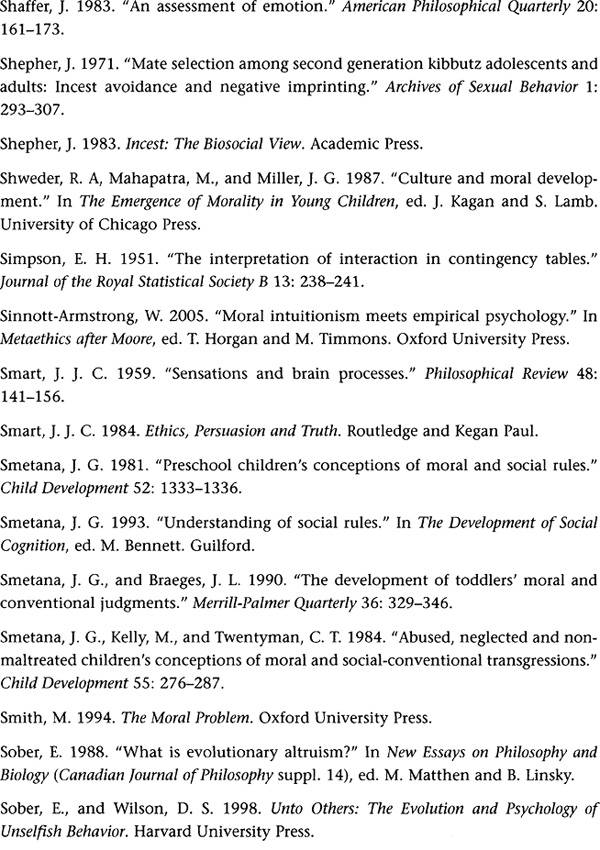


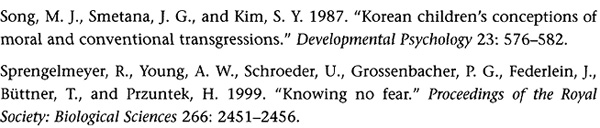


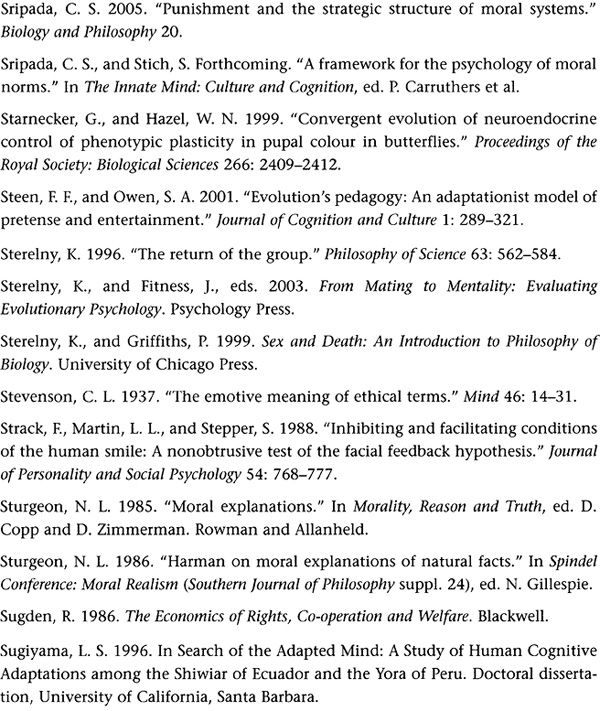


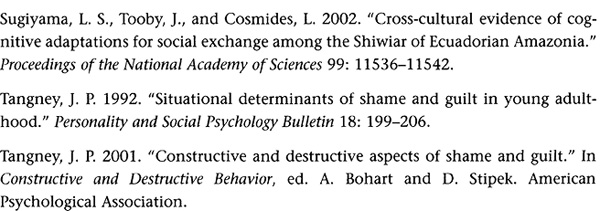


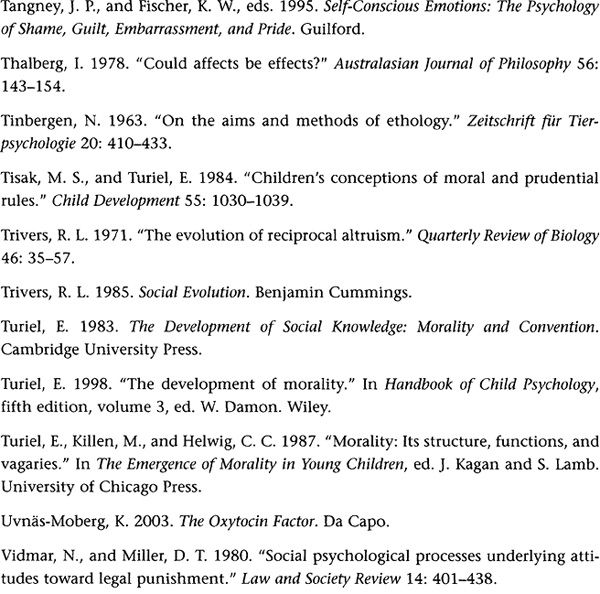


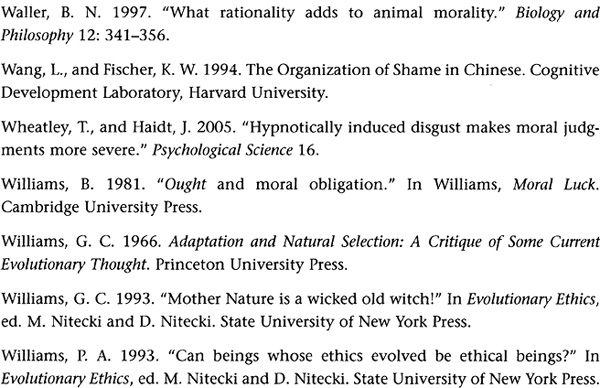


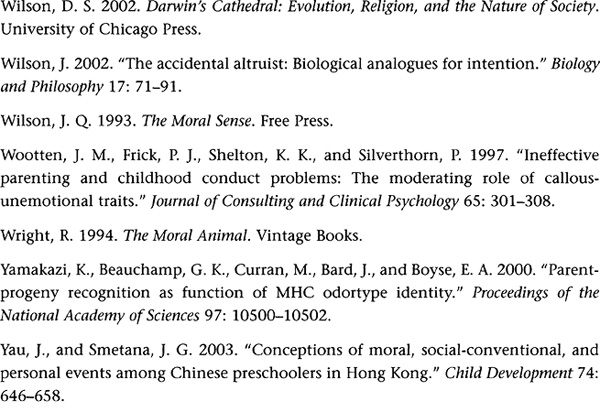


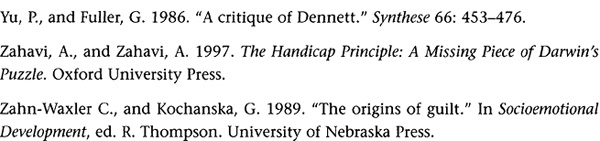












# 注釋

[[1]](#_1_170)這個界定并不是要對先天性這個概念進行分析或一般說明。我并不反對在別的論述中這個術語有不同用法。

[[2]](#_2_165)生物社會學家在某種程度上是咎由自取。菲利普·基徹爾曾經寫道：“生物社會學有兩副面孔。一副是對著非人類動物的社會行為。它目光專注，言論明智，開口說話慎之又慎。而另一副面孔幾乎被擴音器遮住了，帶著極度的興奮，高聲宣講關于人類本性的宏論。”（1985：435）

[[3]](#_3_160)演化心理學家有時會被批評為過分隨便地使用“那個祖先環境”，好像更新世各方面都是始終如一的，并且在演化論上有特殊地位。在很多方面更新世的環境顯然不是始終如一的：人類的一些祖先生活在資源豐裕的地方，而另外一些人則面臨著干旱和貧瘠的環境；對一些人類祖先來說，捕食者構成一個主要問題，對另外的人則不是，諸如此類等。但是，話題要是轉到解剖性狀的演化過程，例如對生拇指和低位喉頭的演化，這些批評者好像就對環境的差異不以為意。人們可以毫無問題地說這些性狀適應于“那個”祖先環境（因為沒有什么東西有絕對適應性）：這個環境在許多方面都有所差別，但同時也展現了足夠寬泛的連續性，因為（比如）對生拇指一般具有普遍的適應性。演化心理學家只不過是對心理機制作出了同樣的斷言。至于“更新世”具有演化的特殊地位：更新世囊括了最近的、時間跨度足夠長的人類演化世系，足以發生重要的生物自然選擇。因此，如果想要理解現在的我們，給更新世以特別的關注是很合適的。毫無疑問，人類身上也有源于上新世、中新世，乃至更早時代的適應（adaptation），但我們所有的適應，包括非常古老的適應，都需要能在更新世維持下來。

[[4]](#_4_158)我并不是主張，依照我之前規定的“先天”的用法，這兩種情況都是“先天的”。但是，在“編寫好”這個詞組的核心意義上，這兩者都是在基因層面上“被編寫好的”。

[[5]](#_5_150)這里有一個真實生活中先天的條件性策略的例子：某些毛毛蟲變成棕色還是綠色的蛹，取決于它們所附著表面的顏色（參見Starnecker and Hazel 1999）。這些毛毛蟲能對它們周圍的環境產生反應，但這種反應是在基因層次編寫好的，以某種的不連續方式對一定的環境變數產生反應。我們可以認為，毛毛蟲的基因組間接編碼了以下條件式反應：“如果遇到X，那就做ϕ；如果遇到Y，那就做Ψ。”把一定數目的毛毛蟲突然放置到沒有綠色事物的環境中，這些毛毛蟲最后變成的蛹可能沒有一只是綠色的。即便如此，變成綠色的蛹仍然是“毛毛蟲本性”的一部分，而變成紫色或檸檬色則不是它們的本性。

[[6]](#_6_144)指1618年至1648年發生在歐洲的一系列戰爭，戰火波及大部分歐洲國家，是歐洲歷史上破壞最嚴重的戰爭之一。——譯注

[[7]](#_7_142)traw-man argument，一種推理和論證的謬誤，指在論辯中先歪曲對方觀點，再攻擊曲解后的觀點。——譯注

[[8]](#_8_142)更為冷靜的謹慎言論可參見Kitcher 1985；Dennett 1995；Sterelny and Griff iths 1999；Sterelny and Fitness 2003。

[[9]](#_9_136)這里借用了Jonathan Dancy的用辭。

[[10]](#_10_134)按照我的定義，助他行為的出現可以是非常偶然的（其實犧牲繁殖適應性的行為也是一樣），但是我也意識到了Jack Wilson（2002）指出的由此產生的困難。

[[11]](#_11_130)比較學究氣的補充：因為繁殖適應性是相對的概念，我們可以合理地宣稱，我們可以通過幫助一個生物的競爭對手來傷害該生物的適應性。無論有沒有被壓扁，這個生物的繁殖能力都為零，但我認為必須承認相對于它競爭對手的繁殖能力，為零的情況可以更好也可以更壞。

[[12]](#_12_128)這種模糊性對應于“因為”這個詞的模糊性。我們可以說，福瑞德關心他的妻子是因為他愛她——這里的“所以”指的是他的動機，是他在慎思中的考慮；我們也可以說，福瑞德愛他的妻子是因為這樣會促進他的基因復制——這里的“因為”指的是因果—歷史性的解釋。這還只是“因為”這個詞諸多截然不同的用法里的兩種，我們一定不能混淆。

[[13]](#_13_124)為了避免閱讀上的冗長乏味，我把具體的例子留給讀者去想象。

[[14]](#_14_120)“即使他們（史前的發明家，即對別人有幫助的人）沒有留下任何孩子，他們所在的部落還是有他們的血親。” ［達爾文（1879）2004：154］。

[[15]](#_15_120)我這樣的表達方式是遵從一個普遍但容易引起誤會的慣例。事實上，同胞兄弟姐妹的基因的相似度高達99%以上，而且任意兩個同種生物的基因相似度跟這也相差無幾。同胞兄弟姐妹之間所謂“50%”的關系（以及表親之間“25%”的關系等）指的是它們因為有共同來源而肯定相同的基因的比例；或者我們也可以把這種比率視為兩者共享任意的等位基因的可能性。

[[16]](#_16_120)我高度簡化了這里的問題；膜翅目昆蟲的基因情況是非常復雜的。

[[17]](#_17_114)要正確地處理親緣選擇，不僅要考慮基因關聯的程度，還必須要考慮受惠者的“繁殖價值”。一個個體與他的祖父母和孫子孫女可能都擁有同樣的基因關聯度（假設每代都有必要的存活時間），但幫助孫子孫女比幫助祖父母對于提高繁殖適應性要有效得多。參見Hughes 1988。

[[18]](#_18_112)Kibbutz（復數形式Kibuttzim），以色列的一種集體社區，傳統上以農業為主。——譯注

[[19]](#_19_113)有人可能更偏向把“互利共生”一詞限制在不同物種成員的關系上，而把“合作”用于描述同一種物種的成員之間的成本—收益關系。這種概念選擇似乎漸漸不流行了，這里也沒有遵循。

[[20]](#_20_111)這個例子似乎表明，“互惠利他主義”不要求在幫助和回報之間有時間延遲。移除寄生蟲的行為既是給寄主大魚報酬同時也是給清潔小魚的報酬。然而令人驚奇的是，在給出這個例子之前的段落里，特里弗斯說互惠利他主義就像是有時間延遲的“共生”現象，并且“時間延遲是這里的重要因素”（1971：39）。關于精確定義“互惠利他主義”的困難的進一步討論，參見Ethology and Sociobiology 9（1988）上的多篇文章。

[[21]](#_21_112)“對稱性”這個詞的意思是每一方所收到的好處都超出其付出的成本。但是在原則上，交往關系中的一方得到的好處有可能遠遠多于另一方。假設B給A4個單位的幫助，這消耗了他100個單位的成本。聽起來似乎是個糟透了的交易吧？不一定：我們還可以假設，A接下來會給予B150個單位的幫助，而這只消耗了她3個單位的成本。盡管交換明顯不平衡，但因為4＞3以及150＞100，雙方都能從中獲益，并且在其他條件都相同的情況下，他們應該繼續維持這個安排。我們通常假定，對于互惠交換而言，重要的是一方只用相對較少的成本就能給別人帶去利益。但這個假設不需要同時適用于交換雙方。在剛才設定的情況里，這個假設對于B來說是不正確的。但是，當這個假設對于其中的一方來說不是正確的時候，為了彌補這一點，就需要這個假設對另一方來說是“極為正確”的：A給出150個單位的幫助，而只付出3個單位的成本。

[[22]](#_22_112)在某些情景中，拒絕幫助他人和懲罰他人之間或許沒有很大的區別，盡管懲罰他人聽起來更為“積極”。例如，如果一群狒狒要終止與群體中的某只狒狒所有來往，這就跟把它殺掉差不多了。這就是我為什么覺得Chandra Sripada的嘗試有問題：他把對遵守道德的兩種解釋，即基于互惠的解釋和基于懲罰的解釋，當作是對立的解釋（2005）。懲罰經常跟互惠相伴隨——特里弗斯在他1971年的文章中就已經提到了。同時也應該注意，“拒絕參與”付出的代價可能跟實施懲罰一樣高昂。如果獅群要是不允許一只想搭便車的母獅子來蹭飯，那么當后者闖進來分享獵物時，它們就會把它趕走，這樣做可能還要冒著受傷的危險。（獅子其實很能夠容忍搭便車的行為；它們的助他行為看起來是受互利共生調控的，而不是互惠，參見Heinsohn與Packer 1995。）

[[23]](#_23_112)可能有人覺得這是在用詞上吹毛求疵。但我見識過太多由這種約定的和半比喻式地使用“利他主義的”和“自私”而導致的混淆，因此，我認為在這里挑剔一些是有正當理由的。

[[24]](#_24_110)嚴格來說，只有重復博弈才需要第二個條件。如果2R比T+S要小，那么在重復博弈中，兩個參與者每人在T和S之間的來回切換比一起合作要更有利。

[[25]](#_25_108)對手若是ALL D（持續背叛的）博弈者，PAVLOV就會很糟糕，一直在合作和背叛之間轉變，輪流地接受P和S的回報（而使用ALL D的對手則獲得P與T）。盡管如此，PAVLOV要是能發現可以跟足夠的其他PAVLOV博弈者互動，它仍然可以離開ALL D的環境。面對另一個PAVLOV對手，PAVLOV只能得到R（不考慮噪聲）。而兩個互為對手的ALL D的博弈者只能獲得P（原因同上）。因此，一個PAVLOV玩家面對兩個對手（一個是PAVLOV博弈者，另一個是ALL D博弈者），他的總體表現要好過一個同樣面對這兩個對手的ALL D博弈者，當且僅當P+S+2R＞T+3P。但是，我們不能說這一定會發生，因為PD模型的標準與P+S+2R＞T+3P以及P+S+2R＜T+3P都是一致的。Nowak與Sigmund更偏向于先依靠某種類似TFT的策略把ALL D“清除出去”，從而讓PAVLOV能夠稱王稱霸。

[[26]](#_26_108)我說的是“大體上解決”，因為達爾文沒有解釋為什么是雌性孔雀成為選擇的一方。答案在于，很多物種的雌性必須投入大量的精力撫育它們的后代，而雄性投入很少也可以應付過去。我相信這個答案最早是被遺傳學家Ronald Fisher ［（1930）1999］理解的。

[[27]](#_27_104)我把性選擇歸到“自然選擇”這個更一般的標題下，而有些學者不這樣歸類。這只是一個使用術語的問題。

[[28]](#_28_104)要是有人反對這里的分類（堅持認為對助他性狀的性選擇不屬于間接互惠的范疇），這對于本書的論述沒有實質的影響。

[[29]](#_29_102)假設助他程度排名從1到12的12個個體要結合成對。每個人都想要和1結成對，除了1本人，他的首選是2。因此（在沒有干擾的情況下）個體1和個體2將會結成對。個體3的首選是個體1，次選是個體2（現在都無法實現），所以他的下一個選擇是個體4。個體4的第一個選擇也是個體1和個體2，但是都無法成功，所以最后滿足于個體3。所以個體3和個體4最后成為一對。剩下的幾對以此類推： 〈5，6〉，〈7，8〉，〈9，10〉，〈11，12〉。（自不必說，要是個體12有個大棒槌，那一切都將不同。）

[[30]](#_30_100)這種說法對文化上的東西是否能被當作復制因子的問題不置可否，因此不能同“模因論”（“meme theory”）混為一談。參見Boyd and Richerson 2000。

[[31]](#_31_96)有人或許會擔心，文化影響的歷史是否長到能夠對人類的基因組構成造成重大影響。但需要記住的是，對于現在的問題（即人的助他性），我們談論的并不是一個全新器官的出現，而是通過親緣選擇和互惠這些過程實現的“微調”。

[[32]](#_32_94)一個人做某事是因為判斷他應該去做，在這種情形下顯然他也是想要做這件事；一個人的義務感會影響他的欲望。

[[33]](#_33_94)“至少”指的是，雖然一個人無法既道歉又不表達悔意，但一個人可以承認后悔但并不因此而做出道歉；除了后悔，道歉還有其他的必要因素，比如承認自己要負責任。參見Kort 1975；Joyce 1999。

[[34]](#_34_92)倘若我把這句話寫成“但我并不是想暗示我有M”或“但是我這么說并不是想表達我有M”，隨后的論證也不會受到影響。

[[35]](#_35_90)C. L. Stevenson，R. M. Hare，P. H. Nowell-Smith及其他學者持此觀點。

[[36]](#_36_90)或許有人會反駁，羅杰的焦慮本身就要納入狐貍獵人的實踐考慮中，無論他們是否在意［對這個論證的精細論述，參見（Nagel 1978）］。但是即便為了論證的目的而允許這一點，仍然不能否定的是，羅杰的道德判斷，“你們的行為在道德上是錯誤的！”這句話旨在傳達的實踐權威性與和“你們的行為讓我痛苦！”類型迥異。如果羅杰試圖在冷靜的爭論中支持他的道德判斷，他不會首先去證明他的痛苦感受。如果要提及任何受傷害的證據，最有可能提及的也是狐貍所遭受的痛苦。

[[37]](#_37_90)嚴格來說，一個命令式必須是祈使句：“關上窗戶”，“不要殺人”，等等。但是，我們按照習慣用法弱化這個要求也沒有關系，比如把“你不應該關上窗戶”與“你不應該殺人”這些句子（分別）當作假言命令與定言命令，盡管它們都屬于陳述句。

[[38]](#_38_87)除了不可逃避性和權威性，伯林克還加上第三種特征：實踐至上性。一條規定具有至上性，當且僅當它總能凌駕于其他類型的規定之上。乍看起來（不陷入康德文本的繁復細節），這三個規范性特征都不能相互推導出來。我這里的論證都沒有涉及至上性。順便也可以提一下，在發表的其他論文里我并沒有使用伯林克的術語，而是在更寬泛直觀的意義上使用“道德權威性”和“道德不可逃避性”這些詞。

[[39]](#_39_83)例如，人們可能認為禁止吃牛肉的規定對印度教徒來說是自我導向的道德規則。但這個假設是站不住腳的，因為我們可以觀察到，印度教徒認為牛是神圣的動物，其福利必須得到尊重。

[[40]](#_40_81)支持這個觀點的例子和實驗來自Haidt et al. 1993。

[[41]](#_41_81)雅諾瑪人是一支生活在南美洲亞馬遜叢林的部落。——譯注

[[42]](#_42_75)皮日波利瓦（Periboriwa）是南美洲神話里的月亮神，傳說月亮神與大地之神結合從而產生了人類。——譯注

[[43]](#_43_75)對于堅持精確的人我有些抱歉，本書的“猩猩”這一術語既指一般的黑猩猩（Pan troglodytes），也指倭黑猩猩（Pan paniscus）。據我所知而言，我的主張非常寬泛，不需要做出這個區分。

[[44]](#_44_73)這里把“should”翻譯成“應該要”，因為漢語里“明天應該下雨”意思不明確，至少沒有明顯的預言含義。更自然的說法是“明天應該要下雨”“明天應該會下雨”。在英語里可以直接只用“should”表達這種含義。——譯注

[[45]](#_45_73)古希臘人非常熱衷討論“沙堆”難題：先讓別人同意“一粒沙子不是一堆沙”和“如果你有一些還不能成堆的沙子，那么你不能通過添加一粒沙令這些沙變成一堆沙”，然后再向她指出，從這兩點可以推出她自己并不相信的結論：十億粒沙不能構成一堆沙子。

[[46]](#_46_71)他在同一頁還寫道：“當我用非常急切的語氣對我的獵犬說（我做過很多次這個實驗）：‘嗨，它在哪里呢？’它馬上就當作一種打獵的信號，并且通常很快速地看周遭一眼，然后沖向最近的灌木叢，嗅嗅有沒有獵物的味道。一無所獲之后，它會抬頭看看附近的樹，尋找上面的松鼠。這些行為不正是顯示了它頭腦中有種一般性的觀念或概念，即要去發現或獵取某些動物？”

[[47]](#_47_69)一般認為，厚重詞匯的特征是它們同時包含描述和評價的成分，因此，在這點上，它們被認為與“單薄的”（“thin”）術語相對立，比如“善的”（“good”），“錯誤的”（“wrong”）。我自己對這種厚重/單薄的區分非常懷疑——我認為，“善的”“錯誤的”（等等）同樣有描述成分（雖然我承認，它們可以有更廣泛的應用范疇）。無論如何，我提到這些術語只是為了方便那些對其比較熟悉的讀者。我的論證完全不取決于厚重/單薄這個區分是否明確或重要。

[[48]](#_48_67)事實上讓事情更加復雜的是，伊尼德之前說西蒙“他這呆瓜笨到都可以算是酷的地步了”。我必須承認，看了這部電影后的幾周內，我在自己的筆記本上涂滿了“呆瓜是一個概念嗎？呆瓜性是一種性質嗎？”這樣的問題。

[[49]](#_49_67)也有人會抱怨我把兩個問題混在一起：一是怎樣才算具有概念X，二是要滿足什么條件才能賦予別人一個內容可以合理地用“X”描述的意向性狀態（比如信念）。但是在這個語境中，我愿意把這兩個問題混在一起，也許是因為當前有關具有概念的哲學文獻紛繁混亂，我才這么大膽。無論如何，很難想象一個人為什么會既承認動物缺乏道德錯誤的概念，但同時又同意動物具有形如“X是道德上錯誤”的信念，或者更寬泛來說，同意動物能夠對道德錯誤做出道德判斷。

[[50]](#_50_67)我把丹尼爾·丹尼特歸為工具主義者主要是因為他經常被這樣看待，并且在很多場合，他聽起來也很像工具主義者。然而我也注意到，丹尼爾·丹尼特拒絕這個標簽，還站在了一個更微妙的立場。因此，我這里對“工具主義者”的說法不能全都明確歸于丹尼特本人。

[[51]](#_51_67)為了簡明起見，我預設了人類最開始的語言是口頭的，但是事實上，有很多其他的合理假說認為最開始的語言包括了肢體動作，話語從肢體動作獨立出來（如果確實從中獨立出來的話）肯定是更晚的事情（Corballis 1999）。即使基于這種假說，3.2節的論證仍然成立。我們需要記住，由于習慣的關系，熟知的肢體動作可以像口頭語那樣清楚地“表達態度”。

[[52]](#_52_67)動物行為學的先驅德斯蒙德·莫里斯在他的暢銷書The Naked Ape（1967）中，推測性地陳述了這個一般假說。這本書引起了很大爭議，甚至在美國最高法院還有過一場關于是否允許小學生接觸這本書的辯論。

[[53]](#_53_65)我在這里參考唐納德·布朗的觀點（2000）。

[[54]](#_54_65)目前有更好的證據證明，覺察他人各種情感表達的能力可以受到選擇性傷害。參見Gray et al. 1997；Sprengelmeyer et al. 1999。

[[55]](#_55_65)如果我們考慮到人們在虛構與其他人的相遇時，會享受和尋求他們通常在現實生活里會避免的情感（恐懼、悲傷等），那么第二個假說比第一個假說會得到更大的支持。這種演化假說認為，虛構和假裝的能力是處理現實生活里的風險和機遇的一種“安全訓練”。自然選擇為了引發這種活動，就讓相應的情感愉快愜意（出于同樣的理由，自然選擇讓吃東西與性行為變得愉快愜意）。參見Steen and Owen 2001。

[[56]](#_56_65)大蘋果（The Big Apple）是紐約市的昵稱。——譯注

[[57]](#_57_63)讀者也許奇怪我為什么不直接說“真誠的言語行為”。但真誠是個更復雜的概念，我在別處有過討論（Joyce，forthcoming a）。

[[58]](#_58_63)Williams 1981對“必須”（must）和“應該”（ought）如何影響實踐慎思做了細致的哲學討論。

[[59]](#_59_63)“我實在不喜歡X”可以是辯護的一部分：“我實在不喜歡X，而且在這種情況中，我用自己的強烈偏好來指導行為的做法是可以接受的。”但顯然，這種辯護的后半部分引入了一個規范性原則。這后半部分通常不會明說：單是“我喜歡咖啡”，看上去就像已經為喝咖啡提供了充分的辯護，但只要與“我喜歡虐待兒童”比較一下，前者的隱含前提（大意即在相關情形中用個人偏好指導行為是正當的）是不言而喻的。

[[60]](#_60_63)托馬斯·莫爾（1478—1535），英格蘭政治家，《烏托邦》作者，因反對英王亨利八世擔任英國國教會領袖而被以叛國罪處死。——譯注

[[61]](#_61_61)此處弗蘭克發展了達爾文在《人類與動物的感情表達》（The Expression of the Emotions in Man and Animals）（1872）中的論證。

[[62]](#_62_61)括號里這句話的目的是將4.2節討論的個人承諾（與本節討論的人際承諾）統一起來。心理學家強調個人承諾，而博弈論學者強調承諾如何影響他人的選擇。（見Nesse 2001：14—15。）就我們的目的而言，這些概念可以有效地統一起來。

[[63]](#_63_61)達馬西奧指出，“他愛說臟話，向人展示自己的不幸，我們據此可以推測他缺乏尷尬意識”（1994：51）。展示不幸指的是Gage自愿在當時某個“怪胎秀”巡演上現身一事。

[[64]](#_64_61)“最后通牒”游戲有兩個人參與（叫他們A和B）。A會領到一筆數額固定的錢，然后必須向B就兩人如何分配這筆錢提出一個方案。如果B接受，錢就按該方案分配，游戲結束；如果B拒絕，兩人都拿不到錢。你可能預期B會接受任何非零數額，而同樣想到這一點A會提議給B最小的非零數額。但有趣的是，這不是真人游戲的結果（Güth and Tietz 1990；Roth 1995）。如果A分給B的份低于某個B視為“公平”的底線，B傾向于拒絕。而A玩家似乎知道這一點——他們分給B玩家的份通常相當可觀。不同的社會似乎有不同的“公平”標準，但A玩家之不只留給B最小非零數額，以及B玩家之不接受最小非零數額，是相當穩定的跨文化性狀（Henrich et al. 2001）。

[[65]](#_65_61)更多關于這個觀點不意味著什么的討論，見Joyce forthcoming b。

[[66]](#_66_61)我不能漏了尼克爾斯接下來的話：“當然，現有證據不足以讓我們對這個假說充滿信心。”（Nichols 2004：176）

[[67]](#_67_59)兩者都是基督教新教群體，且相當有歷史淵源。猶以生活方式傳統樸素聞名。——譯注

[[68]](#_68_59)海特用括號表示“目標單詞”。在這個實驗中，部分被試讀的故事中含有“tries to take”；其他被試的故事則含有“often picks”。

[[69]](#_69_57)這里說的“典型道德判斷”不同于4.2節說的“標準道德判斷”。在4.2節我主張道德判斷通常表達意動狀態，這里的主張是某些道德判斷包含意動狀態的投射。但這兩個意動狀態是不同的。前者是某種類似于接受一個規范性框架的狀態；后者則是諸如憤怒厭惡之類的情感。我并不主張公共道德判斷的功能是表達被投射的情緒；相反，判斷者和他的聽眾通常不會意識到他們的道德判斷是憤怒或厭惡的產物，盡管他們非常清楚在公開褒揚某事時他們是在表達對某個規范性觀點的擁護。（這些主張是作為概念真理或必然真理提出的。）

[[70]](#_70_57)雄性白冠帶鹀只有在聽過典型的“白冠帶鹀鳴”后才能那樣鳴唱（sing）：聽覺隔離的環境中長大的個體永遠學不會正常鳴唱，但只要是聽過正常鳴唱的個體，即便只在雛鳥時略有耳聞，成熟后就能發出完整、復雜的鹀鳴。有趣的是，如果雛鳥聽過兩種不同鳴唱——一種來自成年雄性白冠帶鹀，另一種來自其他鳴禽，那么它只能學會前者。雛鳥需要環境刺激才能學會鳴唱，但它的神經機制會尋找某種特定的刺激，即某種特定的鳴唱，并只對那種刺激產生反應。見Marler and Tamura 1964；Marler 1991。

[[71]](#_71_55)杉山和他的同事認為他們的結果支持一個更具體的假說：人類大腦中存在一個單獨的“作弊偵測子程序”。他們也許是對的（他們的假說甚至也可能是對卡明斯、哈里斯和努涅斯的數據的正確解釋），不過這里我只想強調一個更籠統、更沒有爭議的結果：人類在處理直陳和義務條件句的能力上存在分離。用“作弊偵測”來解釋這種分離的一個問題是，即使實驗使用抽象的、不涉及作弊的義務條件句——比如“如果想做A，必須先滿足條件P”——分離仍然存在（Cheng and Holyoak 1989）。不過，被試者處理“能適”（adaptive）義務條件句的表現確實好過處理抽象義務條件句的表現（Fiddick 2003）。

[[72]](#_72_55)其他研究結果讓人懷疑道義思維不是單一的現象。參見Cosmides and Tooby 1997；Fiddick 2003，2004；Nucci 2001；Smetana 1993。

[[73]](#_73_55)盧梭對這個問題也不能說是毫無知覺。幾句話后他就承認“財產觀念預設了很多其他觀念，人不可能一下子形成財產觀念，而只能逐步獲得。”其實他的論文的一個主要目的就是解釋財產觀念逐步形成的過程。不幸的是他臆測式的解釋（盡管按18世紀的標準來說已是很英勇的嘗試）無助于我們的理解。

[[74]](#_74_55)甚至4.2節關于道德化的思維如何強化社會動機的猜測也不意味著互惠過程就是唯一的解釋。盡管我曾說道德化的思維能讓人更有效地參與互惠關系，那只是道德化思維強化助人行為的例證之一。我的意圖不是說，舍棄眼前小利以求日后大利的助人行為必須滿足互惠關系的成本—收益結構。比如，日后的大利未必是另一個體或一群個體的回饋，而是某種集體行動的結果（換句話說，解釋這種行為的過程是互利共生）。

[[75]](#_75_55)這一段也許僅僅是用詞不當，因為在《人類的由來》的其他地方達爾文都堅決反對心理自我主義。

[[76]](#_76_55)在他1994年的論文（以及1985年的專著第十一章中），基徹爾對描述性的演化倫理學能否有元倫理學的蘊含表示了懷疑。他后來修正了這種觀點（參見Kitcher 1998，2005）。本章的末尾將討論他的立場。

[[77]](#_77_55)丹尼特的話在字面上是正確的：從“是”推論出“應當”，確實經常被稱為“自然主義謬誤”。但他自己隨后就因循了這種用法，從而也就延續了對摩爾的誤讀。

[[78]](#_78_53)以下括號里的頁數都出自這一版本。

[[79]](#_79_53)我得承認，我是不明白摩爾是基于什么原則使用或不適用引號。我這里只是引用了他的標點方式。另外，我也忍不住想指出摩爾的例子的怪異之處：橙子是橙色的，不是黃色的！

[[80]](#_80_51)記住我之前對“給善下定義”的一點不滿。

[[81]](#_81_51)參見普萊斯1956和斯瑪特1959。斯瑪特在他后期的著述中，正確地指出摩爾的語言哲學是“一團漿糊，毫無希望”（1984：22）在身后發表的《倫理學原理》修訂版中，摩爾承認在一些直接跟自然主義謬誤和OQA相關的問題上，他在1903年是混淆不清的。重大的退讓是很明顯的（參見Casebeer 2003：21）。

[[82]](#_82_51)即“馬達加斯加”。——譯注

[[83]](#_83_51)我預設了“……不應當……”是個二元謂詞。就算是多元謂詞也沒有關系，我們可以對例子做相應的修改。（Harman 1975）論證說“……在道德上應當……”雖然外表不然，但其實是個四元謂詞。

[[84]](#_84_51)有些證據表明宗教信念可能也是一種適應（Boyer 2001；D.S. Wilson 2002）。

[[85]](#_85_51)事實上，坎貝爾的非認知主義是否真能滿足他自己描述的標準，還是頗值得懷疑的。這是因為，他把負罪感視為道德的核心，似乎只是將其視為非認知的情緒而已。但我們知道這是錯誤的；負罪感具有相當的認知復雜性（參見3.6、3.7節）。

[[86]](#_86_51)美國學者司馬賀提出的概念。這種決策過程得出的結果不是最優的（optimal），但也在個人和環境的條件限制下，已經足夠滿足需要。——譯注

[[87]](#_87_50)威廉姆斯1985：101。

[[88]](#_88_52)魯什迪是印度裔英國作家，曾因小說《撒旦詩篇》（The Sadanic Verses）在伊斯蘭教徒中引起巨大爭議，伊朗宗教領袖霍梅尼曾因此書下達針對魯什迪的追殺令（fatwā）。——譯注

[[89]](#_89_51)也就是說，對于亞里士多德而言，形容詞“好的”總是作為定語形容詞而不是謂詞性形容詞使用。柏拉圖則持不同意見，他認為，有時候我們使用“好的”意思就是好的，僅此而已，而不是好的什么東西。定語形容詞和謂詞性形容詞的區別可以闡明如下。“方形的”是個定語形容詞，因為我們說X是一本方形的書，意思是指X是一本書以及X是方形的。因此，如果X還有別的什么性質（比如是屬于我的物品），我們就可以說X是屬于我的方形物品。但對于謂詞性的形容詞，比如“大”，就不能這么說了。如果Y是一只大老鼠，我們不能把這拆開成“Y是大的”和“Y是一只老鼠”，因為如果Y還有別的什么性質（比如是我的寵物），這并不意味著它是我的一只大寵物；它有可能是我最小的寵物。

[[90]](#_90_51)基薩拉琴（Kithara）是古希臘的樂器，屬于里拉琴（Lyre）的一種。其名稱或許是“吉他”（“guitar”）的詞源。基薩拉琴師（Kithaorode）是基薩拉琴的演奏者。

[[91]](#_91_51)這句引文出自亞里士多德，Nicomachean Ethics 1.7（1992：13—14）。

[[92]](#_92_51)恩培多克勒認為，四肢和器官最初都是在土地里自發生長出來的，隨后各自游蕩了一段時間（“眼睛獨自游蕩，祈求獲得前額”——Simplicius，Commentary on the Heavens 586.6），最終隨機組合在了一起，組合成各種各樣古怪的混合體。（他特別提到了牛臉人和人臉牛。）“一旦各個部分的組合結果，同要是有目的它們就會獲得的組合結果相一致，這些東西就是被自發地組成了有適應性的形式，從而能夠存活下來；但長成其他樣子的東西已經消亡或將繼續消亡。［亞里士多德，Physics II. 8 （1941：249）］ 令人失望的是，恩培多克勒沒有談論累積演化的機制。

[[93]](#_93_51)凱斯比爾認為道德評價的基礎是自我傷害，尤其見于他對那些漠視他人苦難的人的批評：“……對于周圍人們的需要無動于衷者，在眾多方面都是不正常的：他們無法建立獲取基本生存需要的社會關系，也無法參與豐富多彩的人際交往。”（63—64）

[[94]](#_94_51)在“生物學和倫理學”（“Biology and Ethics”）一文末尾，基徹爾試圖從演化的角度維護道德，而以上也是我對他的主要反駁。后面的6.4節將有進一步的討論。我這里的論證是關于互相沖突的命令（關于“應當”的主張），但要是換成互相沖突的價值，論證也一樣有效力。還需要補充的是，我的論證并沒有預設在正確的道德理論中，道德規范必須能夠壓倒一切其他的考量（我在本書對此不表示任何看法）；我的論證可以依托于一個更模糊但也更少爭議的老生常談：道德規范必須不能被輕易地壓倒推翻。

[[95]](#_95_50)參見Joyce 2000b，2001a，2001b，2003，2004。

[[96]](#_96_50)但這個特例已經足以推翻那條常被稱做“基因謬誤”（“the genetic fallacy”）的原則：知道一個信念的起源從來不能證明該信念是錯誤的。

[[97]](#_97_50)曾經有人向我指出，這個虛構的故事甚至連概念上成立的可能性都不存在。（人們可能會問）一個藥片怎么可能產生關于拿破侖的信念呢？這個問題的答案取決于我們覺得是什么因素決定了信念內容。我承認這個故事確實有可能是不融貫的，但是我的建議是，有些時候，不可能成真的思想實驗也能夠有啟發的價值。

[[98]](#_98_50)這里說你失去了全部有關拿破侖的肯定信念，意味著你依然可以形成諸如“我曾經相信拿破侖在滑鐵盧打了敗仗”和“我好奇拿破侖這家伙是否真的存在過”這類信念——不蘊涵也不預設拿破侖真實存在的信念。

[[99]](#_99_50)做兩點多少有學些究氣的澄清：首先，認為對相關譜系學的知識使得我們的信念不被證成（即我們的信念在此之前還是證成的），和認為這種知識揭示了我們的信念從來都沒有得到證成，這兩者是不同的。更青睞哪個觀點取決于個人接受的認識論理論。可靠主義者（本章后面將會談到）青睞后者。因為我保持中立，所以我的結論在此是模糊的，但我傾向于用前者的方式表達觀點。其次，說一個人持有某個信念是得到證成的，和說一個信念在更一般的意義上得到了證成（或者更好地說，作為該信念對象的那個命題是證成的），這兩者也是不一樣的。這種區別在有些場合是比較重要的，但在此我覺得并不重要。聲明一下，我通篇指的都是前者，但我覺得完全可以改寫成“X的信念p沒有證成”；而如果所有人都同X一樣，我們似乎就有理由說“信念p沒有證成。”

[[100]](#_100_50)有的哲學家懷疑數學命題是否為真，而且數學命題何以為真（即是何種事實使其為真），也是個值得思考的哲學難題。但是我現在討論的這場辯論預設了，如果用來論證道德信念是錯誤或未證成的邏輯，同樣可以用來論證相信1+1=2也是錯誤的或未證成的，那么這種論證就等于是把自身引入荒謬之地。

[[101]](#_101_50)在他1986年的論文里，哈曼把例子降格為一個人用棍子打貓。

[[102]](#_102_50)在討論世界上的心靈屬性時，克倫·本內特（forthcoming）區分了廣義的物理屬性和狹義的物理屬性。這一區分很有益處。狹義的物理屬性是自然科學的法則和概括中用到的屬性（比如：是個電子），而廣義的物理屬性則是由狹義物理屬性以一種可以清晰闡明的方式組成的屬性（比如：我走向廚房的第四步）。非還原主義者否認道德屬性（或心靈屬性，在Bennett的事例中）等同于某種狹義的自然屬性，但不否認其為廣義的自然屬性。在我看來，這兩種等同關系都可以擊退哈曼的挑戰，所以在現在這個場合（但當然不是在所有的場合），我們似乎可以對兩者一視同仁，都歸入“本體論的還原”。

[[103]](#_103_50)本節和下一節的內容使用了Joyce 2001a中2—5章的材料。

[[104]](#_104_50)這里引用的約翰·麥基（John Mackie 1977：35）對于“客觀規定性”之地位的看法。

[[105]](#_105_50)即袋狼（學名：Thylacinus cynocephalus），一種食肉的有袋類動物，學界普遍認為袋狼已于20世紀滅絕。——譯注

[[106]](#_106_50)哈曼（1986：65）也對“公平的觀察者”（“impartial spectator”）理論頗具好感。我不知道這種理論有什么希望能確保實踐權威性，而在這個問題上，它比實踐理性理論前景更為渺茫。一個公平的觀察者的看法，比起醉酒的觀察者或狂熱愛國的觀察者，為什么就更重要呢？雖然有許多人嘗試回答這個問題，但我覺得可以不失偏頗的說，只有原本就同情這種立場的人才能夠信服那些解答。在諸多嘗試中，有一些屬于我在后文將要批評的“自我觀念策略”。

[[107]](#_107_50)再準確一些地說：我們在尋找的，是對于假如一個人做出正確推理，會建議那個現實中的（可能也是推理不正確的）自己去做些什么的一套解釋。參見Smith 1994。

[[108]](#_108_50)科普把這些觀點分別歸于伊曼努爾·康德，克里斯汀·科爾斯嘉德，托馬斯·內格爾和斯蒂芬·達沃爾。

[[109]](#_109_50)對科爾斯嘉德的觀點更細致的批評，參見123—133頁，Joyce 2001a。

[[110]](#_110_46)鮑勃·迪倫，“Highway 61 Revisited”。

[[111]](#_111_46)對于我在4.3節提出的論證，這個自然主義者也得做出同樣的回應，因為沒有權威的道德不會是有效的人際承諾工具。在有風險的合作事業中，下面三種個人，我們會選誰作為伙伴：（A）一個是因為合作會帶來實際好處而愿意合作的人，（B） 一個因為合作是道德要求（不止是因為務實的原因）而愿意合作的人，但同時他也知道，改變自己的欲望就能讓他的道德理由化為烏有，還是（C）一個因為合作是道德要求（不止是因為務實的原因）而愿意合作的人，但同時他也認為，無論他的欲望會怎么改變，他都有好的理由遵循道德？我想，很明顯C會勝出。因此，如果被選中參與這種風險事業是適應的，那么在其他條件都正常的情況下，C類的人比其他人有更高的繁殖適應性。這表明，如果自然選擇對于作為人際承諾工具的道德判斷青眼有加，那它應該更偏愛把道德規定（以某種堅定但可能還不完全成熟的方式）看成是權威和不受普通的慎思標準制約的性狀。這并不必然是在說，人們事實上總是有理由遵循道德的規定，而是說人們傾向于覺得他們有這樣的理由。這個結果就意味著，那種認為不包含實踐影響力的道德自然主義，不太能夠符合我們日常觀念里的道德。

[[112]](#_112_46)有人可能會覺得，這種道德語言的贅余性問題會困擾任何形式的道德自然主義，包括那些想要接納實踐影響力的道德自然主義類型。如果事實如此，對我倒是沒有絲毫影響。但其實對于想要接納實踐影響力的自然主義者來說，這個問題并不是十分棘手，因為他不認為道德話語可以被有關愿望和需求的簡單話語所取代。他想把道德屬性等同于某些（據說是）具有超越欲望的實踐權威的屬性，而且在他看來，對于這些自然屬性的談論仍然可以在個人和人際事務中起到阻止算計和終結對話的理想作用。

[[113]](#_113_46)我在別處論證過更有力的結論：事實上我們應該不相信道德命題（而不是對道德抱著不可知的態度）。我的看法并沒有改變，只不過是在此把那些論證暫時擱置一旁罷了。

[[114]](#_114_46)克勞瑟斯是個認為知識不需要證成的可靠主義者。因此其實他的論證并不是要證明先天信念是證成的，而是要表明它們構成了知識。但是，顯然我們可以對他的論證略加修訂，使其結論成為：先天信念是證成的。我現在正是這么做的。

[[115]](#_115_46)這不能混同于修正信念的融貫主義方法（即反思均衡，“ref lective equilibrium”），該方法最著名的運用見于約翰·羅爾斯的著作。

[[116]](#_116_46)“……要是一個人事先已經真的覺得，是否應當把諸如順應法庭的判決，把無辜之人的處以死刑這樣的行為納入考慮之列，還是個懸而未決的問題——那么我就拒絕跟他辯論：他這個人的心術不正。”（Anscombe 1958：16—17）

[[117]](#_117_46)如果有人要堅持說，不可知論的觀點是我們不知道上帝的存在，那么這就導致了兩種可能的不可知論：有神論的不可知論（即相信上帝存在但承認不知道上帝是否存在）和無神論的不可知論（即不相信上帝的存在但承認不知道上帝是否存在）。這樣一來，我們就需要更多的名號（不同類別的“可知論”），來形容那些有這些信念但同時又宣稱知道上帝是否存在的人。這樣的分類倒并不一定有什么毛病，但我想說的只是，人們通常并不是這樣看待這些立場之間的關系的，我在此也不是這樣看待的。

[[118]](#_118_46)霍爾沃德·利勒哈默爾曾經論證過，從道德的譜系學不能直接推導出道德錯誤論，但在他那里，“錯誤論”的用法是有所限定的，其形式類似無神論。（Lillehammer 2003）

[[119]](#_119_46)我在此寄希望于沒有什么頂尖的網球選手會閱讀本書。

[[120]](#_120_44)該假說的證據來自對于引發惡心的典型事物的考慮。植物常常含有毒素，但相比肉類而言，它們很少攜帶病原體或寄生蟲。因為被感染的肉類通常不會有明顯異樣的味道或口感，食肉者通常要對肉的品質極為敏感。相當肯定的是，證據顯示動物產品遠比植物產品容易引發惡心（Fallon and Rozin 1983）。跨文化的研究也表明，對懷孕早期的女性來說，最有可能感到惡心的食物是肉類。與此發現相符合的是，懷孕期間女性的免疫系統受到了抑制（以便降低把“異物”排出體外的概率），使其更容易受感染或患病。對于個體因免疫系統受到抑制而更易受感染的問題，自然選擇的解決辦法似乎是一種暫時降低對目標產生惡心感的閾值的機制（Fessler 2002）。

[[121]](#_121_44)即globophobia——但后者的含義現在慢慢變成了對全球化的恐懼（這種恐懼倒可能完全理性的狀況）。

[[122]](#_122_44)這位哲學家是柯林·拉德福德（1995）。對拉德福德的反駁論證，參見Joyce 2000a。

[[123]](#_123_44)在討論道德感演化的過程中，達爾文覺得引用馬克·奧勒留（Marcus Aurelius）的說法很是恰當：“這是你習慣的想法，這也將成為你心靈的性格；因為靈魂是會被想法浸染的。”［《沉思錄》，第五卷，引用于Darwin（1879）2004：148］

[[124]](#_124_44)皮浪主義者（該名稱來自Pyrrho of Elea）是全局的懷疑論者，對什么都不表示贊同。

[[125]](#_125_44)據我估計，達爾文應該會很同情彼得·辛格所擁護的那種道德基礎。本書第六章的內容隱含了對于這種觀點的批評。