

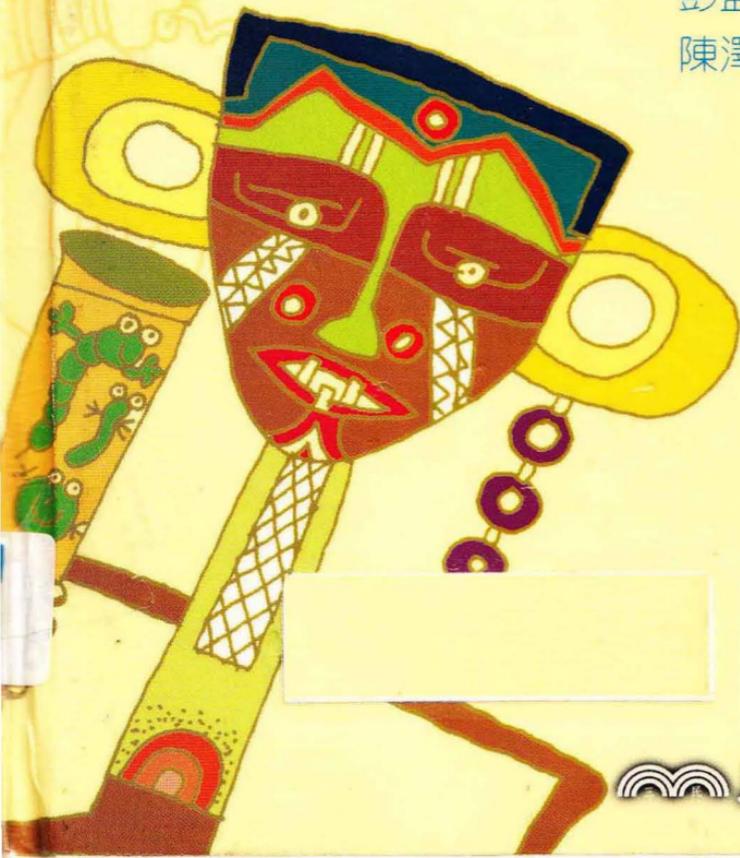
哲學輕鬆讀

# 人心難測

——心與認知的哲學問題

彭孟堯 著

陳澤新 繪



三民書局

# 人心難測

——心與認知的哲學問題

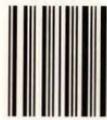
若思考、認知與情緒是大腦的作用，那麼刻骨銘心的愛情與永恆不變的友情，難道只不過是大腦神經系統反應下的產物嗎？這種說法好似貶低了人的價值，卻彷彿同時提供我們關於「人」的另一種想像。身處科技與幻想發達的時代，我們夢想著有一天能夠創造出會思考的機器人——例如擊敗過世界棋王的電腦「深藍」。我們更夢想著有一天機器人能夠更像人：除了思考，還有喜怒哀樂、七情六欲。人類真能辦到嗎？是我們的想像力太過豐富了，還是目前的科技仍不夠發達？本書將最近三十年來哲學家對於這些問題的討論做了一些整理，以淺顯的方式來說明這個發展快速的研究領域裡，重要的哲學問題、哲學學說、及其遭遇的困難。不僅讓讀者對人類價值作深切反思，亦是開拓視野的閱讀經驗。

ISBN 957-14-4287-9

(161)



NT. 200



9 789571 442877

哲學輕鬆讀

# 人心難測

——心與認知的哲學問題

彭孟堯 著

陳澤新 繪



三民書局

國家圖書館出版品預行編目資料

人心難測：心與認知的哲學問題 / 彭孟堯著；陳澤  
新繪。——初版一刷。——臺北市：三民，2005  
面；公分。——(哲學輕鬆讀)

含索引

ISBN 957-14-4287-9 (平裝)

1. 知識論 2. 認知心理學

161

94006345

網路書店位址 <http://www.sanmin.com.tw>

©

人心難測

——心與認知的哲學問題

著者人 彭孟堯

繪者 陳澤新

發行人 劉振強

著作財產權人 三民書局股份有限公司

臺北市復興北路386號

發行所 三民書局股份有限公司

地址 / 臺北市復興北路386號

電話 / (02)25006600

郵撥 / 0009998-5

印刷所 三民書局股份有限公司

門市部 復北店 / 臺北市復興北路386號

重南店 / 臺北市重慶南路一段61號

初版一刷 2005年6月

編號 S 141100

基本定價 肆元

行政院新聞局登記證局版臺業字第〇二〇〇號

有著作權不得侵害

ISBN 957-14-4287-9 (平裝)



## 哲學人的哲學事——序言

### Q 遇見哲學的那天：

第一次接觸哲學應該是大一修「哲學概論」的時候吧！說實在的，那門課無趣得很！我只記得那位老師教到很古老年代的希臘時期哲學時，提到一位哲學家說：「世界是由火構成的」。然後那位老師在黑板上畫了一雙眼睛，眼睛裡有火，在眼睛外面老師又畫了熊熊烈火。然後老師說：「這位哲學家認為，因為世界都是由火構成的，所以我們的眼睛能夠看見東西。」天哪！我在讀大學嗎？

我無意冒犯我的老師。我只是在想，我絕對不會將我一輩子投入在研究這種無聊的問題。或許老天爺覺得我應該學哲學吧！因為系上來了一位大家都很尊崇的老師，我隨著熱潮，修了他開的西洋哲學史。我們讀卡繆的《異鄉人》，幾個同學非常熱烈地討論這本書。我們幾個天真沒開竅的傻小子，一致認為那是哲學，那是我們很想學的東西。就這樣，一路走來，凡是沾上哲學的



課都是我的目標，凡是沾上哲學的書，都是我想讀的。我才不管那是先秦諸子，還是隋唐佛學，我才不管那是柏拉圖的《對話錄》，還是康德的《第一批判》。總之，只要是沾上哲學的，我都想讀一讀，我都想知道書裡面在說什麼。

我一心想出國學哲學的夢想強烈到了極點，可是我剛到國外學哲學時的挫折，也讓我不解到了極點。老師教到一位德國哲學家的學說，提到他關心的哲學問題，改用中文的例子是：「孔明等於孔明」和「孔明等於諸葛亮」這兩句有什麼不同？天哪！我出國留學就是要學這個嗎？哲學太深奧了，我的資質太駑鈍了！我想，在這個知名的大學裡，知名哲學家教的，必定有其道理吧！於是我耐心地一點一點學習，直到取得了博士學位！我逐漸逐漸發現，英美分析哲學真的是我想學的哲學。我很難描述這種思路的轉折。我是在哪一天遇上哲學的呢？在我被他吸引的那一天！

### Q 哲學對作者的意義：

哲學不是坊間的算命，哲學也不用來算命，哲學也不是混淆人心的玄思空想。在分析哲學領域裡學了這麼多年，我逐漸感受到，學哲學不僅讓我的思想更縝密、



分析更犀利，尤其讓我深深體會到哲學與科學那種密不分的關聯。在我看來，哲學跟物理學、數學、生物學和心理學一樣，是一門嚴謹而值得追求的學問，是以理性思維來發掘真象的學問。

每門學問都有其專有名詞，更有其獨特的理論問題。哲學的問題，不論是形上學的、知識論的、還是心與認知哲學的，都是來自於人類純粹的好奇心，想要知道我們是如何獲得知識的、我們的心智能力究竟是什麼能力、我們人類在浩瀚宇宙中究竟有多大的份量…等等。科學家也在探討其中的一些問題。但是我覺得每門科學的背後都有一套哲學想法不容輕視，我更希望瞭解這些哲學的想法。

許多人都說「哲學沒有用！」哲學家除了會用一些讓人退避三分的名詞，說一些讓人聽不懂的大道理之外，似乎沒有任何用處。這是誤會哲學了。如果所謂的「有用」是指能不能讓人賺錢，那麼有太多的學問都是沒有用的。然而如果所謂的「有用」是指對於真象的瞭解，滿足人類天生的求知慾，還有什麼比哲學和科學更有用的？

正如同不是每個人都喜歡物理學一樣，不是每個人都對哲學有興趣，不是每個人對於哲學家提問的問題有興趣。我偏偏碰巧是對於這些問題充滿好奇心的人。我



還稱不上是優秀的哲學家，然而，「舜何人也，于何人也，有為者亦若是！」沈浸在哲學研究裡，是我一生的志趣。

### Q 本書特別之處：

什麼是「思考」？什麼是「認知」？什麼是「情緒」？哲學家發現「人心」真是難以度測。天文學家都已經提出宇宙是誕生於一百五十億年前左右的學說了，我們對於發生在自己身上的情緒、每天都在運作的認知、乃至於每天都面對的自己，卻所知有限！這確實令人詫異！

當代哲學家認為空想是沒有用的，正本清源之道還是在於清楚瞭解「思考」、「認知」、和「情緒」到底是什麼？這門二十世紀新興的心與認知哲學，對於我們的心智與情感展開了一系列的探討。將近百年的哲學研究，雖然還說不上已經徹底瞭解了人類的心智與情感，但經過多年的努力，英美哲學家已經有非常豐碩的成果——至少我是這麼認為。

研究哲學一直給人一種錯覺與挫折：為什麼總是有問不完的哲學問題？為什麼哲學家對於每個哲學問題總是有歧見？為什麼好不容易才有了一點頭緒，覺得某個哲學主張對於問題的解決已經呼之欲出，隨即就有別的哲學家說，那個主張有許多漏洞？這本書介紹的哲學也



不例外。當我們覺得問題即將解決時，就會有哲學家告訴我們還有更艱難的問題在我們眼前。這絕對不是哲學家故弄玄虛，而是既然研究哲學的目的在於追求真象，沒有真正說明真象的哲學學說怎麼能讓我們滿意呢？從正面來看，這其實是好事，每一次的批評與修改就代表往日的跨越了一步。

目前國外已經有不少書籍有系統地整理了這些哲學發展及其成果，唯獨在中文界這類書籍仍然付諸缺如。撰寫本書的目的是在中文界拋出一個引玉的磚石，期待中文界開始熟悉並注意到英美哲學界在這方面的努力與貢獻。

我在本書盡量用淺顯易懂的文字，避免艱澀的學術名詞，來介紹並解說哲學家對於心與認知的研究，因為我將本書的讀者定位在具備一般理解能力的朋友，而不是學術界已有所成的資深前輩。我不知道我成功了沒有。我走的哲學路是曲折的，因為一路上都為艱澀的學術所苦，慨嘆沒有淺顯而適當的指引。我將這本書當作一個開始，期待有更多更多幫助年輕學子的好書出現。

這本書當然沒有辦法將所有這領域的重要問題、概念及學說一一解說，還有許多許多不得不割捨的。在行家眼中，這本書說得太少了。不過，我覺得與其耗費時間去知道所有的學說，不如讓讀者自己有機會思考並發

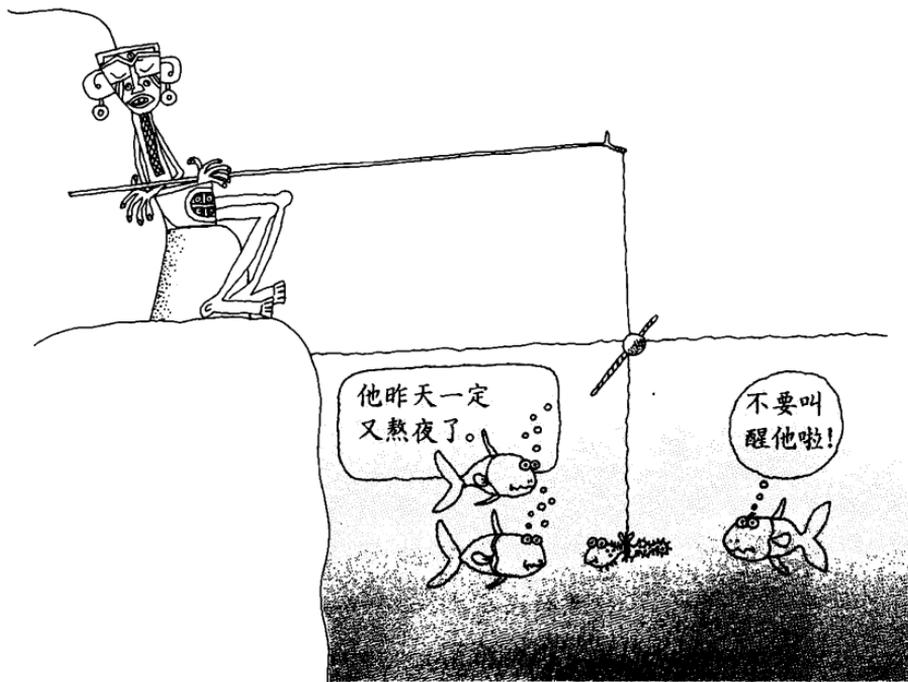


展自己的想法，更有價值。

沒有必要每個人都要作一個哲學家，不過可以肯定的是：每個人都有作哲學家的時刻。偶而，你也在思考這些哲學問題的，不是嗎？希望這本書也能給我的讀者一種感覺、一種想要發掘心與認知真象的衝動！

彭 孟 堯

謹誌於中正大學哲學系



他昨天一定  
又熬夜了。

不要叫  
醒他啦!

# 人心難測——心與認知的哲學問題

目次

## 哲學人的哲學事——序言

### 1

#### 心與認知的哲學困惑

- 人是萬物之靈? ..... 3
- 心理世界與物理世界 ..... 7
- 心存在嗎? ..... 12
- 心與認知的哲學困惑 ..... 19

### 2

#### 心物難題

- 心與物的三個原則 ..... 26
- 科學與因果概念 ..... 29
- 我的心是私密但透明的 ..... 32
- 心的意向性與主體性 ..... 39

### 3

#### 其實我不懂你的心

- 別人也有心智嗎? ..... 50
- 心理因果原則 ..... 54
- 自我類比推論 ..... 58
- 從行為到大腦 ..... 65



## 4

## 行為、人腦、人心

- 科學與物理論..... 70  
行為主義怎麼說? ..... 75  
行為主義的錯誤..... 79  
心與腦..... 86  
大家都錯了! ..... 94

## 5

## 會思考的機器人

- 瘋子的痛與外星人的痛..... 102  
圖林算機..... 110  
人與機器..... 116  
不懂中文的洋人..... 120

## 6

## 人的思維

- 表徵、內容與意向性..... 128  
思維語言假設..... 132  
孔明與諸葛亮..... 136  
雙生地球..... 140

## 7

## 感 質

- 我看到的是綠色的玫瑰..... 151  
中國頭..... 157

沼澤人.....	161
黑白瑪麗.....	166
身為蝙蝠的樣子.....	173

8

動物的心

會算術的馬.....	180
如何問「動物有沒有心智」的問題? ·	186
擬人化的思維.....	189
動物的語言與概念.....	194
欺 騙.....	199

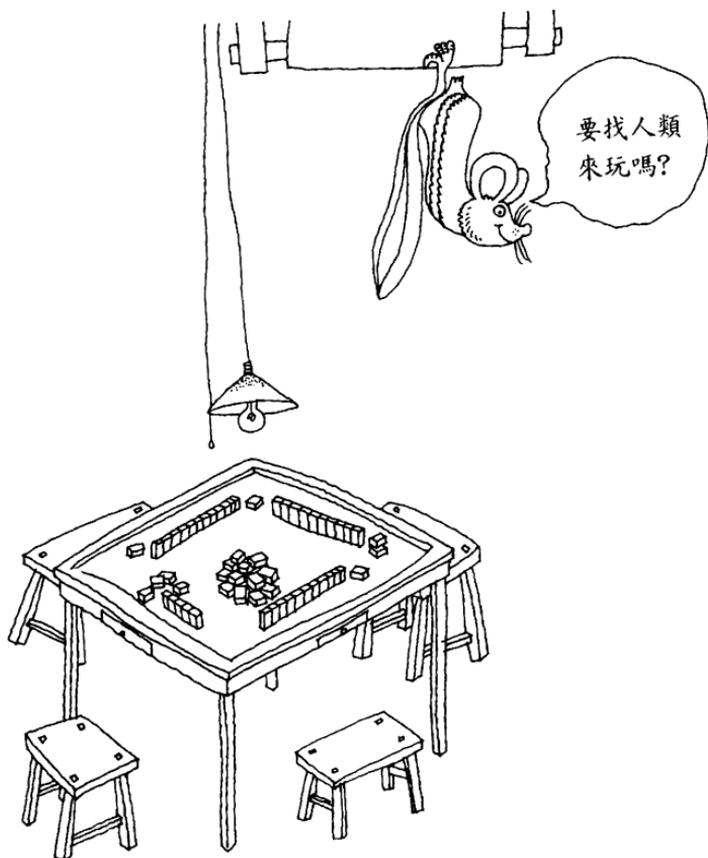
結 論

人心難測? ——心與認知研究的未來·	205
--------------------	-----

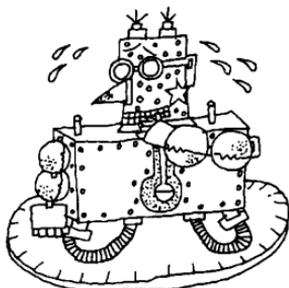
中英名詞索引

1

# 心與認知的哲學困惑



感情豐沛的機器人



唉！

人類根本不懂  
我們的心情。

哼！

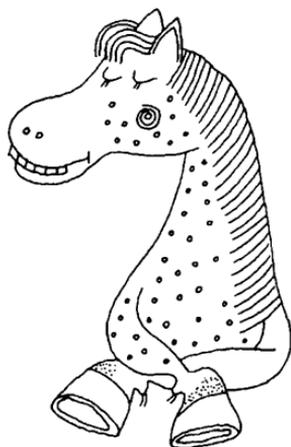
還說我不會算  
數 ...7 ...8 ...9 ...。

NO!

他們只會耍老千！



一身爛泥的沼澤人



數學超強的漢斯公馬



## 人是萬物之靈？

我們常說人是萬物之靈。撇開隱藏在這句話後面的人類傲慢不談，我們人類確實跟其他生物有非常非常大的差異。人跟飛禽走獸的差異究竟是什麼呢？雞貓魚豬有喜怒哀樂的情緒嗎？牠們會嫉妒嗎？牠們會憂慮嗎？牠們會喜歡一些事物、嫌惡另一些事物嗎？許多人都說很多生物具有喜怒哀樂的情緒。即使如此，似乎很難說牠們跟我們人類一樣，會下判斷、會思考、會欺騙、甚至想像力豐富、甚至還會做推論、會自我反省、會規劃

哈哈！別說笑了！  
上帝上次告訴我，  
人類的製作成本最低。  
所以……  
所以祂才做出那麼多個呀！  
哈～～



人生是黑白的黑白瑪麗



未來、會爭功諉過、會找理由來掩飾自己的愚昧。動物好像還沒有聰明到那種程度。

儘管如此，在 20 世紀初，德國有位馬戲團的馴馬師養了一匹名叫「漢斯」的馬。這位馴馬師發現漢斯竟然有計數的能力，而且還會作加減的計算。漢斯這匹馬似乎是很聰明的！如果說漢斯這匹馬沒有心智，人類似乎太傲慢了；然而，如果說漢斯這匹馬乃至於其他的馬是有心智的，會不會又太高估動物了呢？或許我們得好好想一想到底心智能力是什麼樣的一種能力。

姑且不論人類以外的動物是不是有心智，不可否認的，我們人類都有喜怒哀樂、七情六慾等，各種情緒、感受、和情感。尤其我們人類還有很高的心智能力。我們會作判斷、會進行一些推論、會對預定達成的目標作規劃，我們會思考、會想像、會揣測。總之，我們擁有相當高的心智能力。這些心智活動可說是每天都出現、每天都在運作中的。

這些喜怒哀樂的現象、這些心智活動的作用是如此的熟悉，如此的稀鬆平常，以至於我們通常不覺得它們是什麼神秘而值得研究的事情。可是有時候，如果我們注意到這些現象，思考它們的性質，我們往往又覺得困惑。這些喜怒哀樂的情緒、這些思考、推想、判斷的認知能力，究竟是什麼？



其實，我們對於人類這些情緒、感受與認知活動等（統稱為心理現象的事物）所知非常有限。幾千年以來，我們只是將這些現象歸因到神秘不可知的靈魂所具有的力量。隨著當代科學的進步，科學家對於大腦的瞭解愈來愈多後，我們便很快拋棄靈魂的說法，而理所當然地將心理現象當作是大腦的作用，將思考當作是某種腦波，或者某種腦神經活動。

自從電腦出現之後，許多人開始夢想製造一臺會思考的機器人。也有一些人覺得那只不過是幻想，不是夢想。那些用冷冰冰的金屬，加上一些電腦軟體所造出來的「東西」，怎麼可能會思考呢？更別說會跟人類一樣具有喜怒哀樂了。可是，這些反對者是不是太沒想像力了呢？在美國著名的科幻影集《星艦迷航》裡，有個名叫「百科」的人型機器人。他（還是「它」？）擁有跟人一樣的外型，他幾乎無所不知（因為資料庫超級龐大）、他動作迅速、他計算非常神速精準、他能提供具體的建議給艦長作參考、他還能跟別人討論要如何完成所交代的任務。

百科唯一缺乏的是人類所獨有的情感、人的喜怒哀樂、人的愛恨好惡、人的憂鬱悲傷、人的歡欣鼓舞、人的理想熱情、人的卑鄙無恥。即使是影集中的外星人也都有這些情緒和情感。在某一集裡，百科的創造人將情



百科擁有超級龐大的資料庫，但是擁有龐大的資料庫不表示就擁有知識，因為資料本身有真有假，而假資料是不構成知識的！

感晶片植入了他的頭部，讓他短暫擁有了和人一樣的歡樂、喜悅、嫉妒、焦慮與害怕。讀到這裡，你有什麼想法？樂觀的人可能會想：也許等將來科學研究夠發達了，我們就真的能造出這樣一個機器人。也許

等科學更發達了，我們甚至能造出具有跟人一樣的心智能力、一樣能喜怒哀樂愛惡慾的機器人。樂觀的人深信，現在所謂的人工智能(AI)遲早終能實現我們的夢想。

「那只不過是電影罷了！」有人這麼說，因為人的心理現象有一些重要的、根本的性質，是任何精密的機器人都不可能擁有的。機器人怎麼可能像人一樣，歷經刻骨銘心的愛？機器人怎麼可能像人一樣，因為金榜題名而欣喜若狂？像尼采、沙特一樣的機器哲學家，有可能嗎？像莎士比亞一樣的機器文學家，有可能嗎？像李白、杜甫一樣的機器詩人，有可能嗎？這不是科學不夠發達的問題，而是，即使有一天科學都已經完備了，在原則上，科學仍然不能處理心理現象所具有的某些根本特徵。

這些人缺乏想像力嗎？還是這些人對科學太過悲觀了呢？或者是那些樂觀派的想像力太過豐富了？那些樂觀派對於科學過度期望了？其實兩方的說法都過於輕率



武斷了。雖然沒有什麼是不可能發生的，但在此時我們需要的，不是停留於任憑想像力飛揚的階段，因為純粹的可能性，不足以讓我們瞭解我們所提的問題究竟要如何解決。我們需要更多的研究，需要對於心理現象的本質有更多的認識。什麼是思考？什麼是推想？什麼是情緒？什麼是感受？這些心理現象有什麼重要的特徵？這些特徵有可能被科學所解釋嗎？

## 心理世界與物理世界

### 心與認知哲學常用的概念

讓我先引介幾個名詞，以方便我們之後的探討。

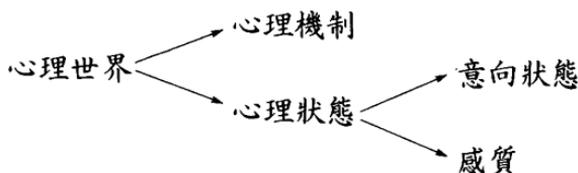
首先，「物」指的是物理世界。物理世界一詞統稱那些原則上可被我們人類觀察的事物和現象，包括我們自己的身體、四肢軀幹和五臟六腑。也包括可被科學研究的事物與現象（包括不可觀察的事物），例如物理學、化學、天文學、地質學、生物學、生理學……等，一般所承認的自然科學所研究的對象。這些事物和它們的性質構成了物理世界。

「心」指的是心理世界。心理世界泛指我們的感官



經驗、痛覺和癢覺、情感、情緒與感受，我們的喜怒哀樂、愛恨好惡、憂鬱悲傷、苦悶憂愁，我們的歡欣鼓舞、欣喜若狂、雀躍不已，我們的相信、慾望、懷疑、想像……等；另外還包括我們的判斷、推想、思考、問題解決、揣測、規劃……等思維活動或心智活動。簡單來說，各種心理現象、心理事件、心理狀態、心理歷程、心理機制等，構成了心理世界（我們暫時不考慮「潛意識」的心理現象）。本書用「心」或「心智」表示整個心理世界；同時，讓我們將具備心智的個體稱為認知主體。

以簡單的方式來分類，心理世界一方面包括了心理機制或歷程，另一方面包括了心理狀態；而後者又再分類為意向狀態以及感質。



- 「心理機制」指的是推論、做決定、規劃、解決問題等等心智能力。舉例來說，你本來想看場電影的，但是你注意到今天是週末，碰巧市區有非常龐大的遊行。於是你改變主意，決定去租片子回家看。你的思考過程就是一個做決定的心理機制。通常心理機制的運作很短暫，



你也很少注意到它的運作。

- 「感質」指的是七情六慾和喜怒哀樂等情感、情緒和感受、知覺經驗、身體感覺（例如痛、癢、喝醉酒的感覺）等心理狀態。在哲學上，還有一種說法：「感質」不是指那些情感、情緒、感受，而是指這類心理狀態呈現給擁有它們的個體的那種方式。例如，你的蛀牙讓你牙痛，你所經歷到的牙痛的那種尖刺感受，稱為你那牙痛的質性特徵，簡稱為「感質」。（在本書中我們不特別對於「感質」一詞的用法作區隔）
- 「意向狀態」指的是相信、慾望、懷疑、判斷、想、規劃……等心理狀態。舉例來說，小明相信小花是女生，小明判斷小花是高中生。這些類型的心理狀態都具有命題內容，因此當代大哲學家羅素稱意向狀態為命題態度。「命題」簡單來說，就是具有真假值（不是真的，就是假的），由中文、英文等各種語言的句子所表達的意義。例如，「宇宙誕生於 150 億年前」這句話所表達的內容就是一個命題。當我說：「我相信宇宙誕生於 150 億年前」的時候，我的意思是，我現在處於「相信」這個心理狀態，而不是「希望」或「懷疑」；而且我所相信的是「宇宙誕生於 150 億年前」這件事，而不是別的事。



羅素 (Bertrand Russell, 1872-1970) 的一生多采多姿，他共結婚四次，因反戰入獄兩次，並獲得諾貝爾文學獎。他是 20 世紀偉大的哲學家。在邏輯、數學哲學和語言哲學等領域有巨大的貢獻。他最重要的哲學研究主要是在 30 歲到 40 歲之間完成的。重要的著作包括：與另一位哲學大師懷海德 (A. N. Whitehead) 合著的《數學原理》、《西洋哲學史》、《人類知識》，以及許多影響當代哲學界的論文。他非常多產，涉獵領域極廣，除了哲學之外，還包括教育、倫理學、政治等領域。

## 心與認知哲學問題的起源

當代心與認知哲學問題是研究心理機制、感質與意向狀態的哲學。這門研究的起源，在我的觀察，大概可以說是笛卡爾「無心」造成的。當然大哲學家柏拉圖和亞理斯多德並不是沒有對人的心智作過討論，只是相對來說，他們在這方面所造成的影響遠遠不及笛卡爾。



論證 (Argument) 是一種推論，它提出一些理由作為前提，來支持自己的主張，或者反對別人的主張。

由於 16 世紀以後，現代科學興起，天文學和物理學都有了令人振奮的發展，科學研究讓人類獲得了愈來愈多的知識、愈來愈多關於物理世界的真象。當時的笛卡爾，便思索著物理世界與我們獲得的知識之間的關



係。也就是說，他研究的首要目的在於知識論。透過「方法論懷疑」進行思考，笛卡爾提出了他的論證：即使一切事物都可被懷疑是虛幻的，但他自己之正在進行懷疑這件事，卻是一件確定的事。試問：任何人有可能對於「自己正在進行懷疑」這件事提出懷疑嗎？試問：任何人有可能對於「自己正在進行懷疑」這件事出錯嗎？進一步來說，當自己正在進行任何心智活動的時候，不可能自己否定自己正在進行的心智活動，因為「自己否定自己正在進行心智活動」本身還是一項心智活動！所以即使可以懷疑、甚至否定一切事物的存在，對於自己的心智活動進行懷疑與否定卻是沒有意義的。於是有一種可能是：只有心理世界存在，而物理世界是不存在的。在笛卡爾這樣的哲學思考裡，使我們走上心智可以獨立於外在事物的立場。也就是「心」與「物」是不同的。從此構成了哲學上的重大主張：心物二元論。

但是令人疑竇的是：心跟物如果是截然不同的，彼此之間怎麼可能有因果交互作用呢？哲學史上有名的心物難題於是產生了。雖然笛卡爾不是以研究心理現象的議題開始，但卻引發了這個激烈的爭議，所以呢，我才說這是他無心造成的！



笛卡爾 (René Descartes, 1596-1650) 法國哲學家、數學家。治學重理性，使用演繹法，嘗曰「我思故我在」。主張心物二元論，開心靈哲學研究的先河。又首創數學座標，發明解析幾何，開數學史的新紀元。主要著作有《方法論》、《沉思錄》及《哲學原理》等。

笛卡爾主張「具有思維性質的心」跟「佔據空間的物」是兩個不同的實體。由於實體是不能生成變化的，而因果交互作用是一種產生變化的過程，因此「心」跟「物」這兩個不同的實體之間不可能有因果交互作用。他的哲學被稱為心物實體二元論。雖然他對於「心物」的說法與當代的認識有相當大的差異，無論如何，笛卡爾哲學的貢獻之一，在於他正視心理世界與物理世界之間的重大差異，並企圖加以說明。

- ◆ 物：佔有空間，具有「擴展性」。
- ◆ 心：不佔有空間，具有「思維性」（至於情緒和感受等所謂的「感覺」，笛卡爾並不重視）。

## 心存在嗎？

當代哲學界已經不再有人提倡笛卡爾的實體二元論。除了由於大多數哲學家不接受笛卡爾的「實體」概念之外，也是由於科學的發達及其研究成果相當豐碩，使得當代哲學家大多走向了唯物論的立場。由於唯物論一詞所承認的是較狹義的「物」，而不能涵蓋「能量」、「力」等，因此大都改用物理論來稱呼這個哲學立場。



本書亦同。當然有些哲學家採取了別的路徑，於是發展出性質二元論或者唯心論的立場。關於各種哲學立場，請參考本節最後所列表一和表二。

這兩張表列出了總共 15 種哲學理論，仍然沒有窮盡所有的哲學立場。我也不打算詳細介紹每個理論，讀者不必被這麼多的理論學說給嚇住了！哲學界的「某某論」不過是指哲學家對於某個主題所提出的比較有系統的言論而已，其中有些主張已經被淘汰了，有些則還在繼續發展。我們要注意的是這些繼續發展的理論！

## 物理論

物理論是種一元論的哲學立場。顧名思義，這立場只承認物理世界的存在。心物難題由於是二元論才有的難題，物理論的立場當然排除了心物難題（也就是，不認為心物難題是個無法解決的困難）。然而，他們如何看待心理世界呢？採取物理論走向的哲學家只有兩種作法。

- 第一種：取消式。逕行否認心理世界的存在，也就是否認人有七情六慾，否認人會推想、計畫與思考。對取消式的物理論來說，心理世界是不存在的。（你能想像嗎？）
- 第二種：化約式（「化約」就是等同的意思）。將心理世



界當作物理世界的一部份，而不是截然不同於物理世界的。對他們來說，心是存在的，只不過任何心理狀態都等同於某種物理狀態而已。

所以我們可以從這兩個強度不同的方式，將各種不同主張的物理論區別為強的、取消式的物理論，以及弱的、化約式的物理論。



邏輯實證論 (Logical Positivism) 源自 1920 年代奧地利一群哲學家 (也是科學家) 的學術聚會，以及德國「經驗哲學學會」的一群學者，在聚會中發展出來的新的哲學走向。在 20 世紀前期，對於哲學、自然科學和社會科學有非常重要的影響。代表人物包括史立克 (F. A. M. Schlick)、卡那普 (R. Carnap)、萊欣巴哈 (H. Reichenbach)、韓培爾 (C. G. Hempel) 等。此學說反對傳統形上學，提倡「檢證原則」作為接受科學的唯一原則。

並不是抱持物理論立場的哲學家，對於如何理解人類心智都有一致的看法；他們之間還是有差異的，也因而發展出不同的物理論學說。以下是對於物理論的簡略說明，到第四章的時候，我們再來做詳細的評估。

### (1) 邏輯行為主義

在 20 世紀初期，由於受到當時心理學主流勢力行為主義，以及哲學界主要勢力邏輯實證論的影響，出現了所謂的邏輯行為主義，是當代最早出現的物理論的哲學立場 (哲學界也有人將這學說稱為「哲學行為主義」或者「分析行為主義」)。這種物理論又可以分為強邏輯行為主義以及弱



邏輯行為主義。強邏輯行為主義完全否認心理世界的存在，而只承認行為和行為傾向而已。弱邏輯行為主義的說法是，任何心理狀態都可以被化約到（等同於）某組行為和行為傾向。至於心理學的行為主義，哲學家則暫不過問。

### (2) 心腦同一論與功能論

許多人認為邏輯行為主義忽略了大腦的作用。在1960年代，心腦同一論應運而生，他們認為心智的運作其實就是大腦的運作而已，任何心理狀態都等同於某種腦神經狀態，心跟腦沒有什麼不同。即使是現在，恐怕仍有多數人是抱持這種想法（或許正在看書的你也是）。可惜的是，這個學說在哲學界只是曇花一現，因為被後來發展出的功能論所替代。功能論認為人心有如軟體，人腦有如執行軟體的硬體。功能論開啟現代人工智能的研究，從此以後，製造一臺會思考的機器人，似乎不再是遙不可及的夢想！功能論雖然不一定採取物理論的立場，卻跟物理論相當契合；功能論雖然不將人心等同於人腦，但是人腦的作用卻在其理論中有一定的地位。

### (3) 取消唯物論

最極端的物理論莫過於完全否認心理世界的存在。強邏輯行為主義當然是這種立場，另外由加州大學聖地



牙哥校區哲學系教授丘曲冷發展的取消唯物論在哲學界也大有名聲。在丘曲冷的哲學裡，我們大家都錯了：我們其實沒有痛覺、沒有歡愉、沒有悲愁、沒有喜悅；我們其實沒有信念、沒有判斷、沒有思想、沒有慾望。他的哲學主張雖然表面看來相當荒謬，倒也不是無的放矢。我們將會在第四章看到他是如何來否認心理世界的。

## 反物理論

並不是所有哲學家都接受物理論，18世紀的英國哲學家柏克萊提出唯心論的學說，認為我們平常所謂的物理現象其實都是心理現象。由於當代已經鮮少有人接受唯心論的思想，本書就不介紹了。

另外還有一些不接受物理論的哲學立場，稱為性質二元論。性質二元論跟笛卡爾的實體二元論不一樣，它不採取「實體」的概念，而從「性質」的角度來瞭解心和物。在哲學史上，平行論、機遇論、交互作用論、以及副現象論等，都屬於性質二元論。這些理論的差別，主要是在它們承認心和物都同時存在，只是對於心物之間的因果交互作用關係，各有不同的說法。由於副現象論在當代已有相當成熟的論證，我會在第七章做比較詳細的說明。



關於心與認知，為什麼會有這麼多的哲學學說呢？其實這些哲學學說的提出，大致來說，有其歷史先後的次序。舊理論對於人類心智的理解遭遇許多困難，迫使新的想法出現。這是很自然的事。在後面幾章，我將只挑選其中幾個跟一般人想法較接近，或者跟當代心與認知哲學的發展，乃至於認知科學的發展相關的理論來說明。

表一：一元論的哲學立場

唯 心 論		
代表人物：柏克萊		
唯 物 論 (物理論)		
學 說	代表人物	簡 述
個例物理論	幾乎所有當代心與認知哲學家	<ul style="list-style-type: none"> <li>任何個例心理狀態存在皆等同於個例物理狀態</li> <li>可與其他物理論和功能論相容 (故與異例一元論不同)</li> </ul>
異例一元論	德威笙	<ul style="list-style-type: none"> <li>不存在任何類型的心理狀態</li> <li>僅承認個例心理狀態存在，且等同於個例腦神經狀態</li> <li>否認任何心理定律</li> </ul>
強邏輯行為主義	維根斯坦 艾爾 史金納	<ul style="list-style-type: none"> <li>不存在心理狀態</li> <li>僅承認行為或行為傾向為心理學的研究對象</li> </ul>
弱邏輯行為主義	行為主義 心理學家	<ul style="list-style-type: none"> <li>承認任何類型心理狀態等同於某類型行為或行為傾向</li> </ul>

心腦同一論 (類型物理論)	史馬特 普雷斯	<ul style="list-style-type: none"> <li>承認任何類型心理狀態等同於某類型腦神經狀態</li> </ul>
取消唯物論	丘曲冷	<ul style="list-style-type: none"> <li>常識心理學是錯誤的</li> <li>不存在任何類型的心理狀態</li> <li>僅承認大腦神經系統為恰當的研究對象</li> </ul>
非化約唯物論	金恩	<ul style="list-style-type: none"> <li>每個類型的心理狀態都隨附於某個物理狀態(腦狀態)</li> </ul>
物 體 論 (與物理論相容)		
學 說	代表人物	簡 述
因果角色功能論	路易斯 阿姆斯壯	<ul style="list-style-type: none"> <li>心理狀態等同於扮演因果角色的物理狀態</li> <li>扮演因果角色的物理狀態隨物種族類而異(在人類就是腦狀態)</li> </ul>
藍紀語句功能論	藍紀	<ul style="list-style-type: none"> <li>心理狀態是由藍紀語句所定義的二階性質</li> </ul>
算機功能論	帕南 佛德	<ul style="list-style-type: none"> <li>每個類型的心理狀態都等同於某類型的計算狀態</li> <li>認知歷程是計算歷程</li> </ul>

表二：二元論的哲學立場

實 體 二 元 論		
代表人物：笛卡爾		
性 質 二 元 論		
學 說	代表人物	簡 述
心物平行	萊布尼茲	<ul style="list-style-type: none"> <li>心理現象與物理現象並無因果交互作</li> </ul>



論		用 <ul style="list-style-type: none"> <li>心理世界的運作跟物理世界的運作是協調的</li> </ul>
機遇論	馬雷布朗赫	<ul style="list-style-type: none"> <li>心理現象與物理現象並無因果交互作用</li> <li>心理世界的運作跟物理世界的運作是協調的，且這種協調來自於神的作用</li> </ul>
雙面向論	史賓諾沙	<ul style="list-style-type: none"> <li>神是唯一的實體，具有心的性質與物的性質</li> </ul>
副現象論	傑克遜	<ul style="list-style-type: none"> <li>心理性質或現象是由大腦運作而產生的</li> <li>心的現象是演化的副產品</li> </ul>

## 心與認知的哲學困惑

由於功能論的出現，當代心與認知的研究，便結合了哲學、認知心理學、語言學、腦神經科學等，以計算和表徵為核心概念，發展出了認知科學。就心與認知的理論研究來看，當代心與認知哲學不僅參與了認知科學的創立，也建構了其核心的理論基礎。當代心與認知哲學家其實同時也是認知科學家。

或許你已經開始樂觀地看待未來，期待心與認知哲學以及認知科學的研究終究能夠徹底瞭解人的心智。只不過，跟幾千年的發展一樣，對於心與認知的哲學研究，總還是有一些揮之不去的哲學疑惑和難題。除了前面提



到的心物難題之外，哲學家遭遇到的第一個困惑是很有名的他心問題。無庸置疑，我是有心智的。但是我憑什麼認為我以外的人也有心智呢？別人有心智嗎？他們雖然外表跟我類似、行為跟我類似，而且似乎也能跟我溝通。可是如此這樣就能肯定他人是有心智的嗎？哲學家指出，這些似乎都還不足以讓我們確定別人跟我一樣也是有心智的。這種否定別人心智的哲學立場看起來相當荒謬！我會在第三章討論這個問題。

除此之外，關於心與認知所造成的哲學困惑，大致上有三類。

## 意向性與心理內容

心理世界有一個很特殊的性質，稱為意向性。稍早提到的意向狀態（或者命題態度），就具備這種特徵。我們可以說，思考的本質就是意向性。當代心與認知哲學從心理內容來理解意向性，而心理內容則是進一步用表徵論來說明。所以，只要能夠從表徵論提出一套適切的理論，來說明心理內容，或許就能說明什麼是意向性。

心理狀態之具有意向性這項特徵，使得物理論的立場面臨重大的困難，因為物理世界的事物並不具備意向性。此外，心理狀態之具有意向性或心理內容，也構成



了發展人工智能的理論障礙。人工智能的基本想法是將人的心智當作電腦軟體一樣，但是電腦所處理的資訊如何能具備意向性或內容呢？這個問題表面來看，似乎不值得一問，電腦所處理的資訊當然具有內容。不過，等我們在第六章對於功能論的基本想法多瞭解一些之後，我們就知道事情不是這麼簡單。

此外，在心理內容的研究中還遇到兩個障礙，難以克服：指涉隱蔽現象，以及雙生地球的假想故事。舉例來說，張三相信諸葛瑾在孫權手下做事，至於孔明的哥哥是不是在劉備手下做事，還是在孫權手下做事，張三就不清楚了。顯然「諸葛瑾在孫權手下做事」以及「孔明的哥哥在孫權的手下做事」這兩個想法在張三心中是不一樣的，張三這兩個想法想的是不一樣的事情。然而諸葛瑾其實正是孔明的哥哥。張三心中這兩個想法是針對同一個人、同一件事的，怎麼可能會是兩個想法，而不是一個想法？我們將在第六章與雙生地球的假想例一併說明。

最後，「意向性」還引起了另一個問題：動物有心智嗎？動物認知行為學對許多類的動物做了大量的觀察研究，似乎有愈來愈多的研究證據顯示，動物是有心智的。那麼，認知動物行為學家是觀察到了什麼呢？究竟我們要怎樣看待動物是否有心智的問題呢？動物是不是有心



智的問題，跟剛剛說到的他心問題，有許多相似的地方。

（我在第八章會做說明）

## 感質與物理論

「感質」指的是感覺經驗、感受、情感、情緒等心理狀態，它們不具有意向性。感質所造成的問題相當多（我會在第七章逐一解說）：

1. 外星人的心智問題：如果有外星人存在，他們會有痛覺嗎？他們也會有情緒反應嗎？
2. 感質對調的問題：設想小王看到紅玫瑰的時候，他腦海裡出現的色彩知覺跟小張看到綠色樹葉時，所產生的色彩知覺，居然是一樣；而且當小王看到綠色樹葉時，所產生的色彩知覺，居然跟小張看到紅玫瑰時，所產生的色彩知覺是一樣的。小王的兩個色彩知覺是對調的！我們能接受這種說法嗎？
3. 感質缺如的問題：機器人有感覺嗎？如果機器人缺乏感質，那麼功能論想必是錯誤的了。然而承認機器人也有感質的存在，卻又使得哲學家有機會指出物理論的錯誤。另外黑白瑪麗也是個重要的例子，哲學家藉此來反對物理論。
4. 無識人的問題：所謂無識人是指生理和行為跟正常人



一模一樣，但是卻缺乏意識現象、沒有感質這種心理狀態的人。無識人的假想故事使得功能論和物理論都必須重新思考心與物的關係。是不是承認感質的存在，有助於我們反省「究竟該不該接受物理論」？

## 意識的主體性

最後的一個哲學困惑，恐怕也是最困難的一個問題，大概是意識的主體性所帶來的。什麼是意識的主體性？舉例來說，我到書店買了一本書，我便是那本書的擁有者。當我將那本書賣給別人，我就不再擁有那本書了。但是，即使換了不同的擁有者，也不表示那本書消失了。所以，我並不是那本書嚴格意義下的擁有者。同樣的，絕大部分的事物即使改變了擁有它的人，它並不會因此消失。絕大部分的事物並不需要嚴格意義的擁有者。

但是意識卻不同。如果有個意識現象發生，則該意識現象必定出現在某人身上。這意識現象若不為此人擁有，也無法轉移或被其他人擁有。「感到高興」這個心理狀態不會憑空出現，必定是有一個人感到高興。感到高興的那個人就是這個心理狀態的擁有者。不但如此，這種擁有是「嚴格」意義下的擁有。我的意識現象只有我自己能夠擁有，我是我意識現象在嚴格意義下的擁有者，



因為我的意識現象不可能變成別人的意識現象。例如，我的高興是我的心理狀態，你雖然能夠意識到你自己的高興，你不可能意識到我的高興，你頂多只能觀察到我高興時的臉部表情、發出的聲音和手舞足蹈的行為而已。簡言之，意識具有主體性。紐約大學哲學系教授納格用身為蝙蝠的樣子來彰顯意識所具有的這種主體性（我會在第七章討論這個性質對於認知科學造成的困難）。

關於心與認知的哲學研究當然還有很多問題，囿於篇幅，無法一一介紹。其實，能夠對於上述這些哲學困惑逐步探索，就已經使得我們對於心與認知有更深的瞭解了。底下就讓我們從古老的心物難題開始吧！

# 2 心物難題



嘿！  
右腳，  
快踢呀！

哼！每次  
都是我！左腳  
呢？你偏心，  
我要罷工，  
罷工……！



嘖！





## 心與物的三個原則

以下三種都是眾所周知的原則，而且它們都讓人相當信服，但是呢，我們將從中找出一個令你意想不到的結論，一起來看看吧！

原則一：

大多數人都相信心跟物有相當大的差別——我們的信念、希望、判斷、我們的喜怒哀樂七情六慾，確實跟物理現象非常不同。有些物理現象是觀察得到的，但是所有心理現象都是觀察不到的。即使你看得到一個人的臉部表情，和他手腳揮舞雀躍的動作，你看不到他內心的「欣喜若狂」那個心理狀態，你觀察不到他內心想要大聲叫的「衝動」那個心理狀態。你頂多觀察到他的行為而已，但行為跟心理狀態是兩種不同的事物。如果你對上述看法也深表贊同的話，我們說你贊同「心物差異原則」(T1)。



## 原則二：

試想，如果你被路邊一塊石頭絆倒，摔了一跤，你的膝蓋撞出一個大洞，想必很痛吧！又例如，風將沙子吹進你的鼻子裡，讓你不自禁地覺得鼻子癢癢的，連打幾個大噴嚏，你能說你沒有不舒服的感覺嗎？如果你的回答都是「是」，你也就肯定了「物理世界的現象導致了心理現象的產生」。再例如，你心裡想著要在空地上種一棵樹，然後你去買了一些樹苗和肥料，一些劇土的工具，後來你就到空地上種了幾棵小樹。這是否也說明你認同「心理世界的現象也會導致物理現象的產生」？總結一下，你認同：心理世界跟物理世界之間具有因果交互作用。我們稱為「心物因果交互作用原則」(T2)。

## 原則三：

最近幾年可說天災不斷。臺灣在 1999 年發生了 921 大地震，印尼前一陣子發生了大海嘯，還有許多地方地震頻傳。古代人或許會說，「這是天神的譴責！」、「這是地牛翻身！」、「這是海神的怒吼！」現代人不再這麼想了。依據現代科學，我們用大陸板塊的移動與撞擊，能量釋放……等方式



來說明地震的原因。對於地震、海嘯等物理現象的解釋，不再訴諸超自然的力量，而是尋求物理的因素。這種說法稱為「物理因果封閉原則」(T3)。對於任何物理現象的解釋，我們都試圖找出其物理的原因。

這三條原則，各自來看，都有其獨立應該被接受的理由，但是同時接受三者卻會產生矛盾。如何產生矛盾呢？先讓我們接受心物因果交互作用原則 (T2)，心理世界跟物理世界之間存在著因果關係。在這同時我們又接受物理因果封閉原則 (T3)，於是這個時候只能有兩個選擇：

- (1)承認心理世界是物理的，而不是非物理的，亦即心理世界其實是物理世界的一部份。
- (2)承認物理世界是心理的，而不是非心理的。

然而，不論哪一種選擇都會違反心物差異原則 (T1)。因此，我們不可能同時接受這三條原則。一方面，每個原則個別來看都有些道理；另一方面，我們不能同時接受這三條原則，這就是心物難題！

照這樣看起來，為了解決心物難題，在上面三個原則當中，我們只有被迫要否認至少其中一條原則。但要否認 (T2) 心跟物之間有因果交互作用的現象，鐵定會



直接面對大多數人立即的反對。所以讓我們暫時別急著否認第二條原則。可是否認物理因果封閉原則 (T3) 似乎會導致我們對當代科學的不信任，甚至可能造成當代科學的崩潰。到底心跟物

的根本差異又在哪裡？否定心物差異原則 (T1) 是否就可解決這個難題呢？

這三條原則究竟要放棄哪一條呢？



- (T1) 心物差異原則
- (T2) 心物因果交相互作用原則
- (T3) 物理因果封閉原則

## 科學與因果概念

依據上一節的介紹，「物理因果封閉原則」是說：如果造成某個事件發生的原因是屬於物理世界的，則這個事件本身也是屬於物理世界的；或者，如果某個事件所造成的結果是屬於物理世界的，則這個事件也是屬於物理世界的。

簡單說，在同一條因果鏈上，只要有一個現象是物理的，則這條因果鏈上所有的現象都是物理的。相對地，在同一條因果鏈上，只要有一個現象不是物理的，則這條因果鏈上所有的現象都不是物理的（如果存在有這種因果鏈）。



(1)當「 $C \rightarrow E$ 」出現，而且  $C$  是物理的，則  $E$  是物理的；而且

$C$ : 原因事件  
 $E$ : 結果事件  
 $\rightarrow$ : 導致

(2)當「 $C \rightarrow E$ 」出現，而且  $E$  是物理的，則  $C$  是物理的。

---

(3)所以，既不可能出現  $C$  是物理的，但  $E$  不是物理的情況；也不可能出現  $E$  是物理的，但  $C$  不是物理的情況。

照這樣看來，我們最好不要輕易否定物理因果封閉原則 (T3)。因為當代自然科學承認因果概念是可以使用的，自然科學可以使用因果概念來描述發生在物理世界裡的一些現象。對於這類物理現象，物理因果原則會主張，它們的原因跟後續的結果都必須是物理的。不過，並非所有發生在物理世界裡的現象都可以使用因果概念來加以描述。例如量子力學大概用不到因果概念。有些現象的發生可能只能用機率的概念來加以描述。還好，這不會影響物理因果封閉原則，只要我們承認確實有些物理現象的發生，是必須使用因果概念來描述的，就足夠了，我們不必考慮其他只能使用機率概念來描述的物理現象。



如果我們否認物理因果封閉原則，我們就可以放棄原先使用的物理因果概念來描述這類物理現象。這表示我們必須容忍這些物理現象有可能是由非物理的原因造成的；這也表示我們必須允許這些物理現象有可能會產生非物理的結果。然而我們一旦承認，出現在物理世界中的因果關係可以有非物理的原因，或者會產生非物理的結果，只怕會摧毀了科學。畢竟，一旦對於出現在物理世界中的因果關係允許非物理的原因或者非物理的結果，我們對於訴諸神鬼、奇蹟、靈異等作為物理現象發生原因的說法，就再也沒有任何理論上的依據來排除了。打雷的原因為什麼不能說是雷公在發脾氣？火山爆發的原因為什麼不能說是火山神在發怒？地震為什麼不能說是地牛在翻身？樹葉落地為什麼不能說是樹葉喜歡飄落的感覺？科學家如果用這種方式來解釋這些物理現象，我們還需要科學嗎？

那麼，我們是不是應該否認物理因果封閉原則？即便是因而否認了自然科學也無所謂嗎？自然科學在當代的成就並不應該輕易否定的，畢竟當代科學的研究成果是如此的豐富，使得我們對於物理世界有非常深廣的認識。在沒有強而有力的理論支持下，貿然否定科學並不是理性的作為；貿然引進非物理的東西或性質來代替科學所提供的物理解釋，並不能使得我們因而更瞭解物理



世界。

我並不是提倡科學萬能的論調，也沒有吹噓科學是唯一提供真理的方法。我只是想：在科學可以對於物理現象提出物理因果解釋的範圍裡，不論是訴諸非物理的原因，還是承認可以產生非物理的結果，都是多餘而沒有必要的，反而更折損了科學解釋的正當性。就人類知識的成長來看，這樣的論調所付出的代價太高，理性上並不值得往這方向發展。所以，在還沒有提出更好的道理之前，我們應該接受物理因果封閉原則（T3）。

### 我的心是私密但透明的

我們應該接受心物差異原則嗎？不容否認，心理世界跟物理世界確實是有差異的；一般人也確實認為心理現象和物理現象是兩種非常不同的現象。心物差異原則就是主張：心理世界是非物理的世界；物理世界是非心理的世界；「心理的」跟「物理的」這兩個概念是互斥的。這條原則蠻符合一般人的認知，可不是？那麼心理世界和物理世界之間的差別究竟有多大呢？哲學家指出，心理現象有以下五項性質是物理現象所沒有的：私密性、透明性、意向性、質性、以及主體性。前面兩項性質是從知識論的角度來說的，後面三項是從存有論的角度來



說的。我先解說前面兩個性質，其餘三項放在下一節說明。



- ◆ 知識論 (Epistemology) 是研究「知識」本質的一門哲學領域，其研究的問題包括：知識的來源、知識的構造、知識的特徵等。
- ◆ 存有論 (Ontology) 是研究「什麼是真實」的一門哲學領域，其研究問題包括：存在、等同、本質、實在等概念。

## 心理世界具備私密性；物理世界不具備私密性

當我很高興的時候，我知道我很高興，而且我之所以知道這點，是直接就知道的，不必做實驗，不必做推論，也不必他人來告訴我。但是你如果要知道我是不是很高興，你就必須要觀察我的行為、我的面部表情、我的語言表達等，然後才能推論出我內心是高興的（這種推論通常不是明顯的、有意的）。所謂心理世界具有私密性指的就是這種現象。

哲學家說，所謂心理世界具有私密性，意思就是說，每個人對於他自己的內心世界具有特許管取的地位。「管取」就是「擷取資訊」的意思。有些資訊是公開的、任何人都可以拿到的。例如，圖書館的書籍報刊都是可被



公開管取的，這些資訊都是開放給公眾的。但是，只有自己才能有直接的管道，來知道自己的內心世界、自己所處意向狀態的內容、自己所處的感質的那種感受；哲學家說，只有自己才具有「特許地位」去取得關於自己內心世界的資訊。

如果張三有可能知道王五所處意向狀態的內容，那必定是由於張三依據王五的外顯行為做了推論或解釋的緣故，或者是由於王五將自己所想的事情說給張三聽的緣故。如果張三有可能知道王五所處感質的那種感受，那必定是由於張三依據王五的外顯行為進行自我類比推論的緣故（關於「自我類比推論」，我會在第三章說明）。我們每個人透過內省的方式，可以直接掌握自己的內心世界，但是別人對於我自己的內心世界是不可能直接掌握的。

相對來看，物理世界是公開的、開放的；物理世界並不具備這種私密性。沒有任何認知主體對於物理世界具有所謂的「特許管取」的地位。物理世界是開放給所有認知主體的，並不侷限在某個認知主體，而且任何對於物理世界所掌握的資訊都是可被傳遞的，可被公開檢測的。凡是我能夠知道的物理現象，原則上你或任何人都可以用相同的方式知道。雖然我的科學知識都是上課讀書學來的，是透過別人才知道的，不是我自己直接知



道的，但原則上科學家所擁有的一切科學知識，只要我受過適當的科學教育，我也能夠以相同的方式知道。所以，物理世界並沒有這種私密性。

「私密性」這個性質所考慮的心物差異是知識論意義上的，因為它所關心的是獲取資訊的方式，對於被認知對象的本質並沒有加以描述。至於為什麼心理世界和物理世界之間存在這種知識論上的差異，主要是從存有論來解釋，必須等到下一節解說後，才比較清楚。

### 心理世界是透明的；物理世界是不透明的

心理世界另一個知識論上的重要特徵，是它滿足透明定理，心理世界是透明的。所謂心理世界滿足透明定理，是指對於心理世界來說，下列兩個推論都是正確的：

(A) 我正處於某心理狀態。

所以，我知道我正處於該心理狀態。

(B) 我相信我正處於某心理狀態。

所以，我正處於該心理狀態。

讓我舉例說明。當我現在很高興的時候，我當然知道我自己現在是很高興的；不可能我現在很高興，而我竟然不知道我自己現在很高興。這是推論 (A) 的情形。另一



方面，如果我真的相信我自己現在很高興，有可能事實上我現在不高興嗎？可能有人會說，也許我現在是自欺欺人，明明不相信自己很高興，我卻不但假裝自己很高興，還假裝相信自己很高興。可是事實上我現在並不高興。不過，若是如此，一般人會說，我其實並沒有真的相信我自己現在很高興。自欺欺人的可能性反而表示：如果我真的相信我很高興而不是自欺欺人，則我真的很高興。所以推論 (B) 是對的。

然而這兩個推論不適用於物理世界。

(錯誤推論) 我正處於某物理狀態。

所以，我知道我正處於該物理狀態。

(錯誤推論) 我相信我正處於某物理狀態。

所以，我正處於該物理狀態。

舉例來說，我們此時血液內的鉀離子濃度是某個指數，但我們通常並不知道。從「我此時血液內的鉀離子濃度是某個指數」，不能正確推論出「我知道我此時血液內的鉀離子濃度是某個指數」，當然也同樣不能正確推論出「我不知道我此時血液內的鉀離子濃度是某個指數」。另一方面，即使我們相信自己能舉重 100 公斤，但未必事實上我們真的能舉重 100 公斤。從「我相信我自己能舉



重 100 公斤」，來推論「我自己能舉重 100 公斤」是不正確的。物理世界並不滿足透明定理。

儘管大部分哲學家主張這兩種推論適用於心理世界，有一些哲學家提出了質疑，因為這些哲學家認為我們人類還有潛意識的心理狀態，或者承認有所謂次識知心理狀態的緣故。什麼是「次識知狀態」？舉例來說，當你在開車的時候，你剛剛遇到綠燈就往前直行，正好有個穿外套的機車騎士闖紅燈跟你擦身而過，差一點就跟你相撞。你一定有踩煞車，你也一定有注意到這輛機車幾乎撞上你。可是你未必注意到那輛機車的騎士有沒有穿外套。儘管如此，你不可能有看到那位騎士卻沒有看到他有穿外套。你之看到他穿外套的視覺經驗稱為「次識知狀態」。

如果將潛意識或者次識知狀態代入上述 (A) 和 (B) 的推論，恐怕會產生不正確的推論。例如，我真的相信我處於某個潛意識的心理狀態，例如那是心理諮商師告訴我的，這不表示我就真的處於那個潛意識心理狀態，因為那位心理諮商師可能弄錯了。另一方面，從我有某個潛意識心理狀態，當然不能正確推論出我相信我處於那個潛意識心理狀態，因為那個心理狀態之所以是潛意識的，就是由於它並沒有浮現出來，被我意識到。

我不是很同意這樣的說法。首先，雖然我們可以接



受次識知狀態，而且認知科學也確實接受次識知狀態，是否要接受所謂的潛意識仍然是可待爭議的。其次，即使承認人類有次識知狀態乃至於潛意識狀態，這並不會影響我所說的論點。因為現在我們所關切的是心與物之間的差異，只要承認至少有一些心理現象滿足透明定理，尤其只要承認所有浮現在意識層面的心理現象都滿足透明定理，而且沒有任何物理現象滿足透明定理，就足以捕捉到心跟物之間的差異了。



語意學：研究語言的意義如何產生、名字或者描述詞如何獲得指涉、意義跟指涉的差異、以及相關問題的一門哲學領域。

必須一提的是，我曾經指出，對於心理世界另有所謂指涉隱蔽的現象，不能跟這裡的透明定理相混淆。心理世界的透明性是一個知識論的特徵；指涉隱蔽則是心理世界的一個語意學（意義理論）的一項

重要特徵。通常來說，每個名字都有指涉，也就是擁有那個名字的那個個體。例如「諸葛亮」這個名字所指涉的是諸葛亮這個人。描述詞也一樣，例如，「美國 2004 年的總統」這個詞所指涉的是小布希。「指涉」這個概念會遇到一個狀況：在意向狀態的脈絡裡，使用名字或者描述詞的時候，它的「指涉」會出現隱蔽的情形。例如，雖然發明紙的人就是蔡倫，從你相信發明紙的人很聰明，



並不能正確推論出你相信蔡倫很聰明。這個指涉隱蔽的現象也是有待心與認知哲學解決的一個重大問題（我在第六章會繼續解說這個問題）。

## 心的意向性與主體性

這一節我們改從形上學的角度來看心與物的差異，包括意向性、質性、和主體性。

### 心理世界具有意向性：物理世界不具有意向性

關於「意向性」在第一章解說「心」一詞的時候，就已經提到了。具備意向性的心理狀態，也就是羅素所調的「命題態度」，都是具有內容的。以下這些都是意向狀態：希望、相信、懷疑、想像、以為、認知、判斷、揣測……等。如果用英文來描述意向狀態，則在「希望」、「相信」等動詞之後，必定緊接著一個名詞子句，用以陳述意向狀態的內容，例如“**I believe that an apple is on the table**”，粗體字部分就是一個名詞子句，用來陳述我們所相信的事情。簡單講，每個意向狀態都「指向」一件事情；那件事情構成它的內容。

舉例來說，當有人希望晚上吃碗牛肉麵的時候，他



所希望的事情，亦即「晚上吃碗牛肉麵」這件事，就是被他所處的「希望」這個意向狀態所指向的對象。尤其，這類心理狀態的「指向」是它必定擁有的特徵。你不可能做一個沒有任何內容的判斷。你不可能一方面在希望，另一方面卻沒有希望任何事情。當你下判斷的時候，你必定是對一件事情作判斷；當你希望的時候，你必定是希望了一件事情。那件事情就是那個意向狀態的內容，就是那個意向狀態所指向的對象。

所有的意向狀態都具備這種「指向一事情」的性質，這種性質就是哲學家布倫他諾和胡塞爾所說的意向性。相對來看，沒有任何物理現象具備這種意向性。天上的彩虹並不會「指向」一事情，更不會必然指向一事情。紅色的玫瑰、茵綠的草地不會指向一事情，更沒有必定指向一事情。或許你會說，紅玫瑰表示愛情，它不就指向「我愛你」這件事嗎？這沒有錯。不過，紅玫瑰本身其實並沒有代表任何事情，是使用紅玫瑰的人，在那種文化下，將紅玫瑰拿來代表愛情的。（60年代的臺灣應該沒有這種作法吧！）所以並不是紅玫瑰本身擁有「指向」這個特徵，而是使用者的內心想法擁有這個性質，並借助紅玫瑰來表示。由於「指向」這種抽象的意向性特徵是心理世界才有，而物理世界絕對沒有的，布倫他諾甚至有如下的名言：意向性乃是心靈的標記。



同樣地，必須順帶提出一點：就如同透明性不是所有心理狀態都具有的性質一樣，並不是所有的心理狀態都具有「意向性」這個性質，並不是所有的心理狀態都具有內容。「感質」這種心理狀態就不具有意向性，也不具有內容。當然，這個情形並不會因此抹煞了心和物之間的差異。

### **質性的感受是心理世界的本質， 但不會是物理世界的本質**

哈佛大學哲學系教授帕南曾經指出，我們也許會有看到一隻粉紅色大象的幻覺，但我們絕不可能會有「痛」的幻覺。就一般的知覺來說，有可能有人真實地看到一隻聖伯納狗，也有可能有人出現看到一隻聖伯納狗的幻覺。在這個例子裡，視覺呈現給張三的表象，跟呈現給王五的表象是一樣的。幻覺之不同於真實知覺不在於那種知覺經驗本身，而在於那種知覺出現的時候，是不是事實上真的有被知覺的事物來造成那種知覺。既然這個世界沒有粉紅色大象，不會有人產生關於粉紅色大象的真實知覺。因此，當有人說他看到粉紅色大象的時候，他一定是在幻覺當中。

但是，我們不可能在沒有痛的情形下幻覺到痛。幻



覺的痛仍然是真正的痛！你可以假裝感覺到痛，然而如果你不是在假裝，如果你是覺得痛，你一定真的覺得痛，那個痛不可能是虛幻的。「幻覺的痛」是一個矛盾的詞。有一些哲學家將物理世界區別為表象與實在（實在界），認為表象並不是真正代表世界的真象。在這種哲學裡，我們只能觀察到事物所呈現的現象，但不可能發現存在於現象之後的實在。幻覺和錯覺更支持了這種哲學立場。因此有一些哲學家認為區別知覺所提供的表象以及表象背後的實在界是有道理的。儘管如此，這個「表象」與「實在」的區別對於感質並不適用。哲學家瑟爾也同樣指出，對感質這類心理狀態來說，表象就是實在。

「表象」與「實在」的區別不適用於感質，正是由於每個感質所呈現給主體的那種質性感受就是它的本質。就像我在第一章用牙痛作例子所說的，當你牙痛的時候，那種尖銳的痛感，就是你那牙痛的質性特徵。這種情形不存在於任何物理世界之中，沒有任何物理的事物和現象具備這種質性的特徵。

### **心理世界具有主體性；物理世界不具有主體性**

感質或者意識現象這一類心理狀態雖然不具有意向性，但它們具有主體性，而且沒有任何物理狀態具備這



種主體性。

「主體性」是什麼呢？「客觀」與「主觀」是我們日常使用相當頻繁的一組概念，這組概念用來描述的對象較常見的包括事實、知識、方法、真理、判斷、立場……等。不過，我們日常使用這兩個概念時，常常有許多混亂和誤解。至於「主體」和「客體」這一組概念，在日常生活比較少見，在學術界才有比較頻繁的使用。即使如此，學術界對於這組概念也常常錯誤地使用。這裡讓我作一些分析，以使得讀者更精確掌握「主體性」的意義。

首先，「客觀」與「主觀」是知識論的區別，這組概念至少可以透過下列四種理解方式來理解：

(1)所謂「客觀」就是沒有成見的、不攙雜個人情緒的、不受到個人偏好左右的。我們平常使用「客觀」一詞大都是這個意義，例如「力求客觀、以客觀的方法、保持客觀的態度、立場客觀、做出客觀的判斷」等，都是在這意義下來使用的「客觀」。相對地，「主觀」指的是帶有成見的、偏頗的、受情緒影響的。

(2)第二種意義的「客觀」是指程序的中立，指的是禁制人為的操控。例如，進行民意調查時，問卷的設計以及問卷的實施對象就已經刻意偏袒某個特定立場、某個既定團體。例如在議事進行時，在程序上就已經有偏



頗的限制。例如在做科學研究時，就已經偏頗於某個學說假設。一個程序之所以是不客觀的，有可能是刻意的，也有可能是無意中造成的，例如性別歧視、種族歧視、年齡歧視、政治意識型態等，都有可能造成程序的不客觀。

(3)「客觀」的第三種意義是指，對於研究對象或研究情境排除人為的介入或干擾。通常這是社會科學研究的一項重要的方法論規範。例如，對於某原始部落的研究，主張研究者的活動不應涉入被研究的族群，更不應該讓被研究的族群發現他們是正在被研究的對象。這個方法論的規範不僅用來約束對於原始部落的研究以及動物的研究，也包括用來約束一般社會科學的研究，例如對於教育情境的研究，研究者出現在教室會影響正在上課的老師和學生的行為表現，以至於有些行為無法被觀察到（不過，近年興起的「質性研究」並不接受這種規範，認為這意義下的「客觀」是不可能的）。

(4)「客觀」的第四種意義則跟所謂的管取有關。凡是資訊可被公開獲得的，稱為「客觀」。物理世界是客觀的，就是因為任何人都可以經由觀察以及適當的科學教育，來獲得關於物理世界的知識。自己的內心世界則是主觀的。我在本章前面說明心物之間的差異時，就已經提到，自己的內心世界對於自己是私密的，只有自己才



具有特許地位去取得關於自己內心世界的資訊。我們之所以將心理世界當作主觀世界就是這個意思。

從這個意義我們可以進一步說明「主體」和「客體」的區別。在這裡，「主體」指的是進行認知的人，「客體」指的是被認知的事物。

除了從以上四種方式區別四種「客觀」與「主觀」的意義，我們還可以從形上學的角度來看這組概念。

(5)第五種意義的「客觀」，意思是說，一件事物的存在或出現與否，跟人的心智活動無關。在這意義下，「客觀世界」一詞指的是人的心理世界以外的世界，也就是物理世界。花草樹木、蟲魚鳥獸、乃至於整個宇宙的一切、以及人的四肢軀幹、五臟六腑等，都是屬於客觀世界。「主觀世界」一詞指的是人的心理世界。所以，像人的喜怒哀樂、七情六慾、人的思考、想像、推理等，都是屬於主觀世界（不過，將人的內心世界稱為「主體世界」可能更為恰當）。我們有時候說：「某人活在他的主觀世界裡」就是這個意思。「主觀」一詞也因而有「不切實際」的衍生意義。

(6)所謂「客觀」，就是真實存在的意思。凡是被承認是真實的事物、現象、或性質都是客觀的。所以，只要你承認物理世界是真實的，物理世界就是客觀世界。由於我們承認人的心理世界是真實的，所以在這個意義下，



心理世界也是客觀的。哲學家將我這個主體所擁有的世界稱為「主體世界」。

所謂「主體性」，就是指自己是一個主體，擁有自己的內心世界。我在第一章就已經說明了這個性質。所謂感質或者意識現象具有主體性，意思是說，如果有個意識現象出現，則必定有主體擁有那個意識現象。我的意識現象只有我能夠擁有，我是我意識現象的嚴格意義下的擁有者，別人不可能擁有我的意識現象。意識現象的存在，跟它的擁有者之間有必然的聯結，它們無法轉移到別人身上。改變了意識現象的擁有者，就等於消滅了原先意識現象的存在，而產生了新的意識現象。例如，我所親身感受到的那種樣子的喜悅如果能夠「轉移」到他人，被別人所擁有，就會變成他人所感受到的他那種樣子的喜悅，而不再是我所感受到的我這種樣子的喜悅。

顯然，物理世界不具備主體性，因為物理世界的存在並不要求一定要有擁有者；即使物理世界的事物有擁有者，換了不同的擁有者也不會影響它們的本質，進而抹煞它們的存在。

以上哲學家分別從知識論和形上學的角度提出了五項心和物之間的重大差異。到此，基於以上種種對於心理世界跟物理世界差異的解說，我們應該可以接受心物



差異原則了。心理世界是非物理的世界；物理世界是非心理的世界。綜合以上所說的，心理世界跟物理世界之間確實具有極大的差異；心理世界跟物理世界之間確實具有因果交互作用；物理因果封閉原則又很難加以否認。個別來看，這三個原則都有相當令人信服的理由，但是我們不能同時接受這三個原則。心物難題能解決嗎？



3 其實我不懂你的心





## 別人也有心智嗎？

我知道我是有心智的，我有喜怒哀樂七情六慾，我會思考、推論、想像。別人呢？別人跟我一樣也有心智嗎？也有喜怒哀樂七情六慾嗎？也會思考、推論、想像嗎？我確實知道我會思考、會規劃未來、我確實知道我有七情六慾、我有情緒、有感受。透過內省，我知道我此時此刻正在想著去玩一趟巴黎，我知道我此時此刻心情愉悅，我知道我此時此刻正品著一杯香醇的清茶，我



知道我此時此刻在享受一下午暖暖的陽光。我還知道我很討厭榴槤的味道，我知道我受不了一些只會問蠢問題的新聞記者，我知道我蔑視反覆無常的政客。總之，我知道我是有心智的，我知道我的內心世界！

那麼，別人也有心智嗎？我能夠知道別人也有心智嗎？我如何知道在我面前的人，跟我一樣，也是有喜怒哀樂和七情六慾？我怎麼知道老陳也會推論、希望、想像和慾求？我怎麼知道計程車司機也是一個會思考、會



已經不痛了，  
可是眼睛有點酸  
酸的……。

推論、會下判斷的個體？我要如何才有可能知道公司的主管是有心智的呢？他人有心智嗎？這就是哲學裡相當有名的「他心問題」。

關於他心問題，哲學家從兩個方向來思考：一方面，他心問題是指，除了我是有心智的以外，有沒有哲學上的道理來承認他人也有心智？另一方面，即使承認我以外的他人也有心智，他心問題是指：我有沒有方法知道我以外的他人也是有心智的？

絕大多數人一定覺得哲學家真是無聊，竟然質疑這種事情，還那麼認真的看待這個問題、那麼嚴肅地思考要如何解決這個問題。「他心問題」居然還會在哲學界這麼有名！表面看來確實是這樣。我只要看看對方的行為，觀察他的臉部表情，聽聽他說些什麼（如果他沒有欺騙我），我不就知道他內心想些什麼了嗎？我不就知道他有



什麼感覺、什麼情緒了嗎？如果有哲學家認為他沒有辦法知道他以外的人是不是也有喜怒哀樂七情六慾，是不是也會思考、會推論、會做判斷，這樣的哲學家未免也太愚蠢了吧！

這個對於哲學家的譴責會不會太過於輕率武斷了？如果你的想像力豐富一點，科幻電影看多一點，或許你會遇到這種說法：也許我以外的人都不是真正的人，而是外星人或者科學家用超級電腦製造出來的、外型跟我一樣的精密機器人。這些機器人都是外星人或者科學家用某種儀器操控的。他們的操控非常精密、非常細緻、非常周延，以至於我以外的「人」的行為都跟真正的人一樣，跟我自己一樣。我自己根本無法辨別我以外的那個「人」究竟是不是跟我一樣，真的有心智。

「電影看多了！」你大概會這樣抱怨吧！訴諸科幻電影般的情節看似玄想空談，其實是可以在哲學上進一步深入發揮，來申論他心問題的。不過，我不想用玄想的故事來說服你。讓我們換一個方向來思考。讓我們再想想看，我們平常到底是怎麼知道自己以外的人也是有心智的。照剛剛的說法，我觀察到他人的行為，包括他的臉部表情和他發出的聲音，然後我認為他之出現這類行為，一定是由於他處於某種心理狀態的緣故。這是一種從行為反推內心世界的推論。



仔細分析，我們發現在進行反推的過程當中，我們用到了哲學家所說的心理因果原則。同時我們還用到了另外一種推論，稱為類比推論，也就是將對方的「心理一行為」的因果關係類比到我自己的情形。我之所以做出這類行為，是由於我處於這種內心狀態的緣故；所以你如果做出跟我類似的行為，必定是由於你也跟我一樣，處於相同的內心狀態；如果他做出跟我一樣的行為，必定是由於他也跟我一樣，處於相同的內心狀態（暫時假設你和他的這些行為不是模仿，也不是偽裝）。

然而哲學家所說的「他心問題」就是企圖指出：心理因果原則無法用來解決他心問題，自我類比推論在這個問題的解決上也非常薄弱。哲學家認為我們沒有任何可靠的、令人滿意的方法或原則，來知道我以外的人是不是也有心，是不是也有喜怒哀樂七情六慾，是不是也會思考、推論、做判斷。我其實不懂你的心！我其實不可能知道你的心！

底下讓我逐一解說什麼是心理因果原則，以及自我類比推論是如何進行的，以便於思考哲學家所說的他心問題有沒有道理。



## 心理因果原則

也許有人認為哲學家的他心問題難以說服一般人。畢竟，主張在講臺上講得口沫橫飛、興高采烈的老師其實是沒有心智的，認為在旁邊嘮嘮叨叨的父母其實是沒有心智的，認為那個口口聲聲愛我的情人其實是沒有心智的，實在不是一件能讓人信服的事。我雖然不能內省到別人的心，但是觀察對方的面部表情，聽他說出有意義的話，而不是發出一堆無意義的雜音，觀察他的肢體動作，讀他寫的日記和自傳等，就可以知道對方有哪些心理狀態、有哪些想法、有哪些意圖、有哪些慾求，就可以知道對方確實是有心智的主體。這是因為對於他人心智的知識是建立在「心理—行為」的因果推論之上。這個推論用到了心理因果原則：

任何個体外顯行為的肇因是他所具有的內心狀態，而且這種「心理—行為」之間的因果關係具有規律性與普遍性。

依據這個原則，我們應該可以找出一些心理因果定律或通則。舉例來說，假設沒有其他因素干擾，如果你記得你曾經跟朋友約定在火車站出口見面，而且你希望



準時赴約以免讓朋友著急，你必定會做出準時赴約的動作，像是提前出門、搭計程車前往目的地、避開經常塞車的地方……等等。又例如，當你覺得口渴的時候，你想買點飲料來喝，你就會走到有賣飲料的地方，去買杯飲料。使用心理因果定律或通則，就可以進行「心理一行為」的因果推論（通常這種推論不是刻意進行的，而是自然地、不知不覺地進行的）。

心理因果推論就是說，如果我觀察到有人做出了某類的行為，我就能反推這必定是由於那個人具有某類心理狀態的緣故。所以，只要觀察在我面前的人出現的行為，我就可以經由反推而知道他是具有心智的人。簡單說，接受心理因果原則，建立各種心理因果定律，就可以藉著從果推因的推論，知道他人具有心智，進而解決了他心問題。

使用心理因果推論的首要條件，當然是心理因果定律的建立。心理因果定律的建立有兩種管道：透過認知科學對於視覺、語言習得、推想以及各種認知機制的研究，來建立心理學的通則。

建立心理因果定律的另一種管道，是透過一般對於「心理一行為」因果關係的觀察，經由歸納而建立起心理現象與個體行為之間規律的通則。這些構成了我們一般大眾生活的常識，所以哲學家稱之為老百姓心理學，



又稱為常識心理學。常識心理學主要是針對認知主體所相信的事、所慾求或希望的事，以及主要由這兩類心理狀態所導致的行為，來建立心理因果定律。一般人作心理因果推論時，所依據的是由常識心理學建立的心理因果定律。

「心理一行為」的因果推論以及心理因果原則是大多數人都接受的。可惜的是，有什麼理由接受心理因果原則呢？換個方式來問：有什麼理由承認任何兩個人只要具有相同或類似的心理狀態，就會產生相同的或者類似的行為？稍微反省一下，我們應該可以發現，相同或類似的心理狀態，在不同的情形，不同的人，是會產生不同行為的。例如，張三和王五都相信待會就要下雨了。但是他們一個決定打傘，一個決定讓雨淋個痛快。另一方面，不同的心理狀態也常有導致相同行為的情形。例如，張三之皺眉頭，是因為他為了幾分鐘前剛被公司裁員而苦惱不堪；王五之皺眉頭，則是由於他想模仿張三的樣子；陳六之皺眉頭，則是因為他對於王五的動作表示不滿。

這個觀察所凸顯的是：心理因果原則所主張的那種規律性，似乎並不完全適用於所有的心理因果現象。不但如此，恐怕真正的行為與心理狀態之間的關係不是一對一的，而是多對多的。換句話說，當某個環境刺激出



現時，個體會同時有多個心理狀態出現，並進而產生多個行為。很少有人是僅僅處於一個心理狀態，產生一個行為。因此，心理因果原則是不是恰當，不無疑慮。

除此之外，雖然借用心理因果原則來推論對於他人心智的知識，表面來看似乎很有道理，但是在哲學思考上恐怕有些問題：我們會犯下循環論證的錯誤。別忘了我們的問題是：有沒有任何原則或方法可以讓我知道我以外的人是有心智的？然而在我使用心理因果推論的過程當中，我必須先預設心理因果原則是適用於對方的，然後我才能夠依據我對於對方行為的觀察，來

進行從果到因的推論，也就是從他的行為反推導致他行為的內心狀態是什麼。但是，這表示我必須先確知心理因果原則是適用於對方的，也就是說，我必須先承認對方確實是具有心智的個體。如果我不能確認這一點，我憑什麼認為可以將心理因果原則套用到我以外的他人呢？他心問題不正是在詢問我如何能正確知道這點嗎！

讓我們反問自己：我們為什麼不對公園裡的那塊大石頭採用心理因果原則？我們為什麼不對牆壁上的掛鐘



循環論證 (Circular Argument) 一種推論的謬誤：如果有人問你，你為何主張某件事？你提出理由來辯護，可是在你給的理由當中，竟然出現了你原本要辯護的主張。你的辯護就被稱為「循環論證」。通常出現這種錯誤都是無意的，而是由於思考不周的緣故。



採用心理因果原則？我們為什麼不對阿米巴原蟲採用心理因果原則？如果有人說公園裡的那塊大石頭一直不動，是因為那石頭正在那個地方思考，為什麼我們會認為這只是一句開玩笑的話，鐵定不會將這話當真？但是假想公園裡有個人一動也不動地呆呆站在樹下，而有人說他正在思考。他的說法卻不一定是開玩笑的話。

在日常生活中，我們是由於先承認自己以外的他人是具有心智的主體，才會對他人採用心理因果原則，依據對他行為的觀察來進行「心理—行為」的因果推論。所以心理因果推論的使用已經預設我們承認對方是具有心智的主體。訴諸心理因果推論無法解決他心問題。

## 自我類比推論

有些哲學家認為心理因果推論雖然失敗，但是如果稍加修改這個原則，並且配合類比論證，或許可以解決他心問題。這些哲學家的想法是：第一步先對我自己做觀察，將心理因果原則套用到我自己的情況，描述介於我的心智和我的行為之間的因果關係。第二步再將我的情形類比到別人身上。

我將心理因果原則套用到我自己的情況，來發現類似下面的規律：



在某類狀況裡，暫時假設沒有其他干擾因素，某類環境刺激的出現，會導致我處於某類心理狀態；而後我之處於該類心理狀態，會導致我出現某類行為。

例如，觀察我自己之後，我發現我如果跟朋友有約，如果沒有其他干擾因素，我會試圖準時赴約；如果我被石頭（或任何堅硬的重物）砸了腳一下，我會感到腳痛，還會皺皺眉頭、唉喲兩聲。我可以繼續發現各種環境刺激、我的心理狀態、以及我的行為之間的規律。

接著，我觀察到別人跟我在相同狀況下，遭遇相同環境刺激的時候，別人出現了跟我相同或者類似的行為。我採取類比的推論方法，推論出別人在那情境下之所以產生跟我相同或者類似的行為，是由於他跟我處於相同的心理狀態、他跟我進行相同的認知活動的緣故。以下就是使用心理因果原則來進行的自我類比推論：



(前提一) 在某類情境下，我面臨某類環境刺激  $S_1$  出現時，我之出現  $B_1$  類行為是肇因於我處於某類內心狀態  $M_1$ 。

(前提二) 在相同情境下，我觀察到他人在面臨環境刺激  $S_1$  出現時，那個人會出現  $B_1$  類的行為。

===== [ r ]

(結論) 那個人處於  $M_1$  類內心狀態。

其中，[ r ] 表示在所有前提都為真的情形下，結論為真的機率數值，這是因為類比論證是一種歸納推論。

同樣的類比推論架構可以用在其他的「心理一行為」的因果情境。在另外一種情境下，如果我面臨另一類環境刺激  $S_2$ ，我會出現另一類心理狀態  $M_2$ ，並進而導致我產生某類行為  $B_2$ ；那麼當別人處於相同情境，面臨相同環境刺激，也出現跟我相同的那類行為時，他想必也是處於跟我一樣的一組心理狀態。

我還是用剛剛的例子來說明。在觀察我自己之後，我發現我如果跟朋友有約，如果沒有其他干擾因素，我會試圖準時赴約。類比來看，在相同情境下，如果你也跟朋友有約，你也會試圖準時赴約的，只要沒有其他因



素干擾你的行為。所以如果我發現你確實有約朋友，你也確實準時赴約，而且沒有其他干擾因素，我應該推論出，你一定心中也想準時赴約。我還可以將關於我準時赴約的例子類比到其他人的情形。又例如，在觀察我自己之後，我發現如果我被石頭（或任何堅硬的重物）砸了腳一下，我會感到腳痛，還會皺皺眉頭、唉喲兩聲……等。所以在我觀察到你也被石頭（或任何堅硬的重物）砸了腳一下，而且你也皺皺眉頭、唉喲兩聲……等，我就可以反推你一定也是感到腳痛。不但是你，我這個被石頭（或任何堅硬的重物）砸到腳的例子也可以類比到任何其他他人。

心理因果原則搭配類比論證，看起來好像比較有說服力了。其實不然！這個自我類比推論僅僅建立在針對我自己的單一一個個體的情形來進行類比的。這樣的類比太過於薄弱了！試想，如果你在某大學只看到一個學生開車上學，你就推論其他任何一個你立刻會看到的學生也是開車上學的，這樣的推論合理嗎？只有在隨機採取足夠的樣本下進行推論，才是可接受的合理推論。

不但如此，這種自我類比推論還有一個困難。我上一節才提到，在我將心理因果原則套用在他人身上時，我就已經預設對方是有心智的了。即使對於我自己套用心理因果原則是無庸置疑的，有什麼理由主張，對於我



以外的其他個體來說，介於環境刺激和個體行為之間，一定還有個心智在作用呢？有沒有可能其他的個體其實只是由無線操控的一堆「破銅爛鐵」而已？會不會對方只是一個不具生命、沒有心智的生化機器人而已？

我們當然不是試圖從有這種可能性，就進而斷言我們沒有辦法知道他人的心智。僅僅想像一些可能性當然不足以造成哲學困難。提出這些想像的可能性，其用意跟之前的說法一樣：我們沒有任何方法或原則，來排除對方其實是沒有任何心智的個體這種可能性。例如對方只不過是一個由無線操控的不具生命、沒有心智的生化機器人。這是因為我們不知道心理因果推論適用於哪些個體，所以我們不知道可以針對哪些個體來進行自我類比推論。除非我們先行認定對方是有心智的，否則我們怎麼知道要對哪些個體使用心理因果推論以及自我類比推論呢？但是，我們原來的問題不正是：我怎麼知道我以外的任何人是有心智的呢？

到目前為止，我們的嘗試都失敗了。難道我們真的就此承認我們不可能知道別人的心智嗎？當代知識論大師雷勒並沒有這麼快就放棄。他認為這種搭配因果原則，以進行自我類比的推論，還是可以獲得合理支持的，而其中的關鍵則是訴諸所謂的最佳解釋推論。

假設你曾經被蜜蜂狠狠叮咬一口，之後你便感覺疼



痛無比，並因而產生了某些行為、某種臉部表情、還有哀嚎。現在你觀察到張三在類似情形下，被同種類的蜜蜂狠狠叮了一口，張三也表現出跟你類似的行為、類似的臉部表情和哀嚎。如果你不是使用「張三感覺到痛」來解釋他的哀嚎，你要如何解釋他的哀嚎？

也許你會說，張三沒有感覺到痛，他只是在演戲、或者他只是在刻意模仿你的行為而已。不容否認，有時候這解釋是對的。不過，你要對你以外的所有人的所有行為，都做這種解釋嗎？

除了這個問題之外，雷勒指出，在你認定張三並沒有感覺到疼痛的情形下，就算你還能解釋為什麼張三會出現那些行為、那種臉部表情、和那幾聲哀嚎，你仍然必須考慮下列兩種情形：

(甲) 雖然張三表現出那些行為、臉部表情和哀嚎，張三沒有感覺疼痛。但是或許除了張三以外，其他絕大部分人在被蜜蜂叮一口之後，不但表現出那些行為、臉部表情和哀嚎，他們跟你一樣，是會感覺到疼痛的，而不是跟張三一樣，沒有感到疼痛的。只有很少數的一些人才是像張三那樣，不會有痛覺的；絕大多數人是跟你一樣，會有痛覺的。

(乙) 不僅是張三，所有你以外的其他人在接受那類刺激後，雖然都表現出那些行為、臉部表情和哀嚎，



但是都不會感到疼痛。

雷勒指出，不論是（甲）的情形還是（乙）的情形，你都必須提出更完整的解釋。在（甲）的情形裡，你必須解釋為什麼有些人（雖然很少數）做出了那些行為，可是還是沒有那痛覺，而另外有大部分人卻比較「正常」。為什麼張三和少數人跟你和絕大多數人如此不同？在（乙）的情形裡，你必須解釋為什麼只有你才有那痛覺，而所有你以外的人雖然出現那類行為，卻沒有那痛覺？為什麼只有你是如此獨特？

總結來說，如果你假設張三乃至於你以外的所有人是沒有心智的，你不但要另行解釋張三的行為，你還多了這兩種情形必須要解釋。看來，否定別人的心智，似乎反而增加你不少額外的負擔。所以雷勒認為，與其去假設張三沒有心智，乃至於你以外的所有人都沒有心智，還不如假設你以外的所有人都有心智更好。承認他人有心智，然後用他的心理狀態來解釋他人的行為，比起否認他人有心智，然後再設法解釋他人的行為，這兩種解釋似乎第一種比第二種要好。

到此，許多人認為：心理因果原則、自我類比推論、再加上最佳解釋推論，這三件共同使用，應該能夠解決他心問題。讓我接下來對於他心問題再作另外兩點思考。



## 從行為到大腦

他心問題其實還涉及了要如何瞭解「行為」的問題。在自我類比推論中，一個重要的條件是：在接受相同或類似的環境刺激下，別人表現出跟我相同或類似的行為。但是，如何去說兩個行為是相同或者類似的呢？如果只是純粹從肢體動作（包括發出的聲音）來看待行為，我們比較容易斷定兩個行為是不是相同的（相似的）。但是，純粹從肢體動作來看待行為，似乎還不足以讓我們掌握心智與行為之間的關係。

其實我們對於個體行為的瞭解，同時就已經涉及到如何詮釋這些行為：我們會賦予這些行為一些內容，而不是單單從生理的、物理的角度來看待行為。同樣皺眉頭的動作，在我是因為我在深思一個艱難的哲學問題，在他是因為他欠了一堆不知怎麼還的債。在開會的時候，當你看到別人把右手舉起來的時候，你不是僅僅從生理學或者物理學的角度來瞭解，你還將他這個動作當作是他想要表達意見的行為，或者當作他想要表示贊同（或者反對）的行為。同樣這個舉起右手的肢體動作如果是出現在教室裡，除了上述的可能之外，有可能是學生被點名的反應，也還有可能是他在搗蛋。所以，我們似乎



不能那麼簡單地將「行為」當作肢體動作。

那麼，到底「行為」是什麼？如果「行為」不只是肢體動作、不只是臉部肌肉糾結一團、不只是發出某些聲音，那麼「行為」是什麼？似乎「行為」是在生理、物理的動作之上還另外加上某些特徵才行。這特徵初步來說，是指「行為」是有內容的，「行為」反映了一個人在做出某個肢體動作或者發出某些聲音時，代表了那個人產生的某些內心想法或情緒。

我之前曾經提到，心理因果原則的一個問題是，除非我已經確定他人是有心智的，否則我沒有理由對他人套用「心理—行為」的因果定律。現在我們看到心理因果原則的另一個問題：要能夠對我以外的他人套用「心理—行為」的因果定律或通則，我必須先對於他人所做出的行為所代表的意義有所瞭解，但是要瞭解他人每一項行為的意義，似乎只有在對他人套用心理因果定律後才能發現。若是如此，我等於已經先承認（先預設）他人具有心智了。因此，從觀察一個人的行為來反推他的內心世界，解決不了他心問題。

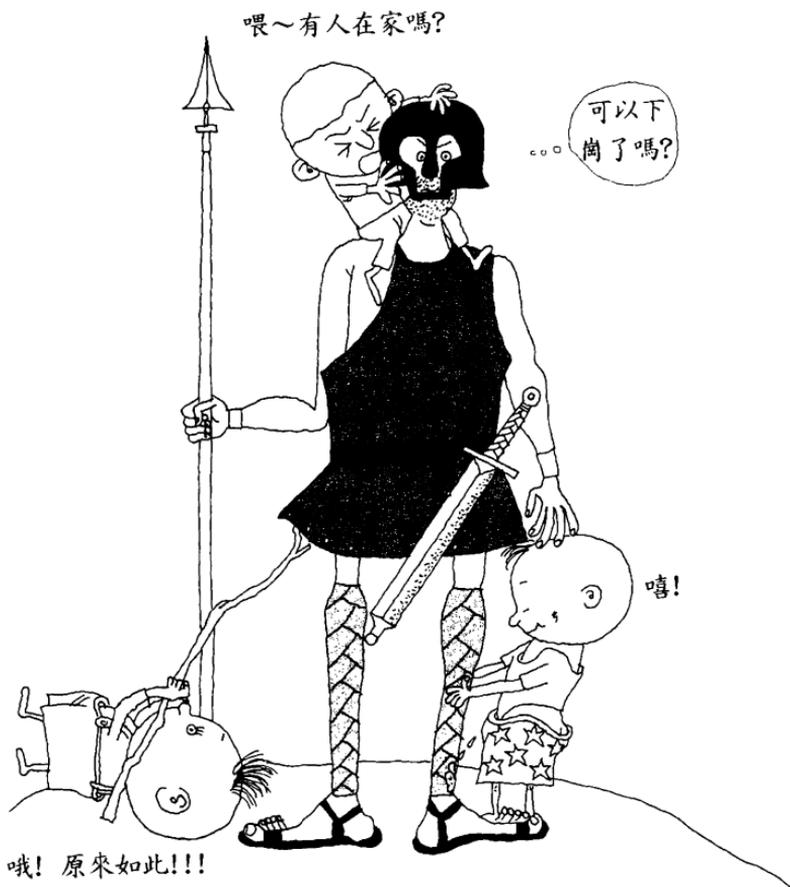
或許，他心問題沒有真正獲得解決的一個關鍵，可能是我們忽略了大腦的作用。介於心理現象跟行為表現之間，應該還有大腦神經系統的作用。畢竟，不僅肢體動作與臉部表情的發生，跟大腦的作用有密切關聯，心



理現象跟大腦神經系統是息息相關的，這一點也應該是無庸置疑的。在承認心理現象跟行為之間是有因果關係的同時，似乎也應該一併考慮大腦神經系統的作用。如何將行為、心智、與大腦神經系統三者放到適當的理論位置呢？讓我們繼續探究這個問題。



# 4 行為、人腦、人心





## 科學與物理論

物理論在西方哲學界有相當悠久的歷史。古希臘就有哲學家提出這個世界是由當時所謂的「原子」構成的。至於他們是不是同時也徹底主張心是由原子構成的，並不是很清楚。大概他們當時還沒有想得這麼遠吧！在西方近代哲學史上，有位哲學家拉美特利提出「人即機器」的主張，是比較清楚的一種物理論。機器的運作，在當時是純粹的機械作用。例如，手錶的運作可以用彈簧和齒輪等零件的運轉來說明，而不需訴諸靈魂、神力等所謂「超自然力」。

這種機械觀在近代是非常盛行的，隱藏在機械觀背後的，就是物理論的哲學立場。機械觀不僅用來說明機器的運作原理，也用來說明各種物理現象，包括天上星體的運行。甚至於原本是訴諸靈魂來解釋的心理現象，也完全改由機械觀的方式來說明。擁有心智並不是什麼神奇的天賦，擁有心智也並不因此讓人跟手錶有什麼重大的差別。

在近代時期，科學的昌盛，讓世人對於科學有極高的信心，不僅認為遲早科學可以解開所有的謎，掌握世界的真象，也認為科學所掌握的就是世界的一切。對於



人類心智的瞭解也不例外，遲早對於心智的研究會構成科學的一部份。西方傳統以來，對於心智的瞭解，僅止於靠內省的方式來做研究。但是在當時這種宰制的機械觀、科學研究的發達、以及對科學的高度推崇與信心的時代，期望對於人類心智也能進行科學的研究，是很自然的事。

在 19 世紀後期，德國的心理學家馮德在這種科學樂觀的氛圍下，首先開展了所謂的「實驗心理學」，可說是科學心理學的濫觴。蘇俄的心理學家巴夫洛夫做了著名的實驗，首先開發了從個體的行為來研究認知；美國心理學家華生則提出了如何研究人類心智的「科學」方法，並且給了這套方法一套理論基礎，這是 20 世紀宰制了心理學研究大半個世紀的行為主義，美國的史金納更是心理學界赫赫有名的行為主義大師。

行為主義心理學對於什麼是科學的想法，決定了它

史金納 (B. F. Skinner, 1904-1990) 當代行為主義大師。主張心理學對於心智的研究不應該採用內省的方法，而是要以外顯的行為作為研究的題材，探討環境刺激與行為反應之間的關係。他從蘇俄心理學家巴夫洛夫的「古典制約學說」發展出所謂的「工具制約」的行為主義心理學，乃心理學三大勢力之一，對於心理學與社會科學的影響無遠弗屆，直至 20 世紀 70 年代才逐漸消褪。





如何研究人類心智的方法。所以我得先簡單說明當時的科學觀。

當時的哲學家和科學家認為科學所研究的對象必須是可觀察、可度量的；其研究方法講求客觀，藉由可操控的因素，來進行可公開的、可重複的實驗，以找出現象之間的關聯，然後核驗假設，進而提出理論以解釋和預測現象。相較於以內省的方法進行研究、對內省內容進行報導與分析的研究方式來看，將行為當作心理學的研究材料，似乎是科學多了、客觀多了。這種對於科學是什麼的想法，到現在仍然影響著大多數人；這種對於科學的高度讚嘆與堅強的信心，到現在仍然深植人心。很多所謂的「學問」為了爭取一般人的信任，總是會試圖以科學的身份出現：做了很多實驗，找了很多證據，使用很多昂貴的儀器，建立了很多令人眼花撩亂的數據。

行為主義主張，心理學如果要成為一門真正的科學，在對於人類心智進行科學探討時，只能以「行為」作為研究的題材，因為行為是公開的、是可以度量的、可觀察的、可重複的。在行為主義眼中，心理學的研究要建立的是環境刺激與個體行為反應之間有規律的聯結。不可觀察、不可度量的任何現象都不是科學研究的題材。所以行為主義認為內省法並不是心理學恰當的研究方法。



行為主義心理學的想法在當時的哲學界也引起了回響，並獲得了哲學學說的支持。在 20 世紀前半時期，邏輯實證論的科學觀籠罩著學術界，社會科學界對於什麼是科學的想法，基本上仍然不脫離邏輯實證論的主張，心理學也不例外。邏輯實證論對於當代西方哲學的影響更是鉅大。限於篇幅，讓我用非常簡短的方式作一些介紹。

邏輯實證論是對於 19 世紀哲學的一種反動思潮。當時有一群哲學家提倡：任何主張或者陳述都必須具有認知的意義，否則我們在理性上不可以接受它。所謂「認知意義」就是指一個陳述不是為真，就是為假。而且一個陳述究竟是真還是假，是有原則來決定的，這個原則就是邏輯實證論非常有名的、也是在其理論中最核心的可驗證原則。什麼是可驗證原則呢？簡單來說，如果一個陳述可以在邏輯上推論出觀察語句，這個陳述就具有認知意義，邏輯實證論將這種陳述稱為「可驗證的」。「觀察語句」是一種句子，它描述了我們使用五官知覺所能觀察的事物和現象。舉例來說，「水是透明的」就是一個觀察語句。

所以，像「上帝是完美的」這種句子就不具備認知意義，因為它是一個無法被驗證的陳述，從這個句子沒有辦法推論出任何觀察語句。邏輯實證論所提出的可驗



證原則，還可以幫助我們區別真科學和假科學。例如，風水學、占星術等，雖然偽裝成科學的面貌，其實都不是科學，都是假科學。它們之所以都是假科學，是由於它們的內容都不可能滿足可驗證原則。

你大概很快就會想到，在物理學裡，有許多句子提到類似「電子」、「磁場」、「夸克」等，不可能被人類觀察到的事物和現象。難道邏輯實證論也要否定這些科學陳述嗎？讓我們將這類型的陳述稱為「理論語句」。邏輯實證論同時有兩種說法：第一種是認為在科學理論中出現的理論語句確實沒有認知意義，



理論化約 (Theoretical Reduction) 是一種介於兩個理論之間的邏輯關係。傳統上，所謂將甲理論化約到乙理論，就是從乙理論邏輯上推論出甲理論的意思。當然，這種推論不是直接的，還必須建立「橋樑」以溝通兩個理論才行。不過，這個傳統的說法在 20 世紀後期已經站不住腳。新的說法是使用「印射」的觀念來說明理論化約的意思。

其實沒有「電子」、「夸克」這些東西，對於這些東西的科學陳述既不是真，也不是假；理論語句只是方便解說可觀察現象的工具而已。

第二種說法是採取所謂理論化約的主張，將使用到理論語句的科學說法，全部化約到只使用觀察語句的說法。如果這是做得到的，在科學裡出現理論語句，來描述不可觀察的事物或現象，仍然是可以接受的作法。

邏輯實證論的說法如何影響心



理學呢？由於心理現象是不可能被觀察的，心理學對於心理現象的陳述也是一種理論語句。因此，按照邏輯實證論的說法，這些關於心理現象的陳述都只是工具而已，方便我們解說所觀察到的行為。從這一點來看，行為主義以個體的行為作為心理學的研究材料，顯得理所當然。尤其，「行為」是物理的現象、是可觀察、可度量的，心理學只有研究「行為」才是真的科學。

## 行為主義怎麼說？

行為主義其實應該算作是一個思潮，因為它深深影響社會科學的研究，也深深影響我們對於心智的看法。哲學家分別從方法論的角度和存有論的角度，區別了(1)方法論行為主義（又稱為「心理學行為主義」），以及(2)邏輯行為主義（又稱為「哲學行為主義」或者「分析行為主義」）。方法論行為主義是關於心理學研究方法的一種主張。在方法論上，心理學的研究應該從考量環境刺激與個體行為反應之間的關係著手。至於介於環境刺激和行為反應之間的「黑盒子」，心理學不必過問。對行為





主義來說，心理學所要觀察的是：環境刺激與行為反應之間的聯結型態、聯結的規律性、行為反應的類型、行為反應的頻率、行為反應的強度……等。

介於環境刺激和行為反應之間的黑盒子會是什麼呢？黑盒子可能是人的大腦神經系統、可能是人的心智、也可能是靈魂。但是行為主義認為，不論黑盒子是什麼，心理學都不必加以理會，因為只有環境刺激和行為反應才是可觀察、可度量的，介於其間的「黑盒子」不論是什麼，都是不可觀察、不可度量的。

方法論行為主義雖然對心理學和社會科學影響重大，畢竟不是哲學家主要關注的課題。哲學家關心的是：究竟有沒有心理世界？如果人真的有心智，我們要如何說明心智的本質？行為主義的哲學立場稱為「邏輯行為主義」。邏輯行為主義既然是一種哲學主張，不會涉及心

理學如何研究人類心智的方法問題。它的主要人物是維根斯坦和艾爾，尤其艾爾在1949年發表了《心的概念》一書，詳細闡述了邏輯行為主義對於人類心智的立場。在他的學說裡，心物難題是虛幻的，因為犯了他所謂的範疇謬誤。



所謂「方法」跟知識的獲得有關；所謂「方法論」則是對於知識（理論、假設）以及依方法而採集的證據之間的哲學關係，進行探討的一門哲學領域，屬於知識論的一部份。



「心」跟「物」分屬不同的範疇，沒有所謂因果交互作用的問題。尤其他更進一步主張，人的心智之相對於人的身體，就好像機器中的鬼魂一樣，都是虛幻的。我們不會用機器裡有鬼魂來解釋機器為什麼會動，為什麼對於人的身體的運作，我們要訴諸心智的作用呢？這是沒有必要的。

邏輯行為主義依其立場的強弱可以分為：強邏輯行為主義以及弱邏輯行為主義。「強」與「弱」的差別在於強的理論取消或者否認了心理世界，弱的理論則沒有這樣極端，而是採取化約（也就是等同）的觀念來思考。這種「取消」和「等同」所造成的理論差別，除了造成邏輯行為主義有強弱之別外，我們很快就會看到，在其他派別的物理論裡，「取消」和「等同」也會造成類似的理論差別。

強邏輯行為主義主張：這個世界根本沒有所謂的心理現象。艾爾所說「機器中的鬼魂」就有這個意思。讓我舉科學史上的一件事來做一個類比。19世紀的物理學家認為光的傳播是需要媒介的，它的媒介稱為「以太」。不過，隨著科學的進展，已經沒有科學家承認有以太這種東西了。事實上本來就沒有「以太」這種東西，是19世紀的科學家誤以為有罷了！同樣地，否定心理世界的哲學家會說：事實上人類是沒有心智的，只是長期以來，



我們人類一直誤以為自己有心智罷了！

這種哲學立場當然是非常極端的。相對來看，弱邏輯行為主義的主張就溫和多了：任何類型的心理狀態其實只不過等同於某類型行為或行為傾向而已。例如，「痛」這種心理狀態等同於某種類型的行為和行為傾向，例如皺眉頭、唉喲幾聲；「高興」等同於另一類型的行為和行為傾向……等（至於哪一類型的行為，則需要行為心理學家去找出來）。

這種「等同」的關係跟所謂「水等同於具備  $H_2O$  分子結構的物質」中的「等同」意思完全一樣。所以心理狀態跟行為之間不會是因果關係，這是由於因果關係通常被認為是介於兩個不同事物之間的關係。不但如此，由於這是一種等同關係，就跟我們不會因為「水等同於  $H_2O$ 」就否認水的存在一樣，弱邏輯行為主義並沒有否認心理世界的存在。

原則上，方法論行為主義跟邏輯行為主義是各自獨立的。也就是說，原則上可以允許同時抱持方法論行為主義，但是不接受邏輯行為主義的哲學立場。例如現今人工智能所謂的圖林測試（見第五章），是方法論行為主義下的作法。但是人工智能的哲學立場則是反對邏輯行為主義的。



## 行為主義的錯誤

行為主義在心理學界佔有很大的勢力，它的影響直到 70 年代才消退。行為主義在哲學界則面臨哲學家提出的許多反對意見，在 50 年代後期就被其它物理論的哲學立場所取代。對於後續的哲學家來說，邏輯行為主義並不是一個能說明心的本質的理論，僅僅以行為或行為傾向並不足以說明什麼是心智。當然，心理學行為主義的失敗，不表示邏輯行為主義的錯誤；而邏輯行為主義的錯誤，也不足以表示方法論行為主義是錯誤的。這是必須留意的。

帕南 (Hilary Putnam) 在賓州大學與喬姆斯基 (N. Chomsky) 研習語言學。後來他一方面在哈佛大學師從蒯因 (W. V. Quine, 當代哲學大師)、王浩 (當代最重要的華裔哲學家)、路易斯 (C. I. Lewis, 當代模態邏輯重要創始人、知識論學者與價值哲學家) 和懷特 (A. White, 知識論學者)；另一方面又在德國師從萊欣巴哈 (H. Reichenbach, 當代科學哲學大師)。



帕南是當代哲學大師，其研究領域廣泛，涵蓋數理邏輯、數學哲學、邏輯哲學、語言哲學、心與認知哲學等，且都具有重大的貢獻。他在語言哲學最著名的莫過於「雙生地球」的假想故事，他的「因果指涉理論」與另一位語言哲學家構成了主流思想的意義理論。在關於心智的研究方面，他是算機功能論的主要倡導人。



對於邏輯行為主義的批評很多，我這裡介紹幾個比較關鍵性的批評。首先，哈佛大學哲學系教授帕南提出了兩個很有名的駁論來批評邏輯行為主義：完美偽裝者論證以及超級斯巴達人論證。

## 完美偽裝者

你一定看過很多的電影和電視影集。對於每年度許多頒獎典禮也不陌生，大家最關注的大概是奧斯卡金像獎最佳演員吧！獲獎的演員被認可具有高超的演技，能完全融入劇情的表演，彷彿演員自己就是劇中人物一般。然而演戲畢竟只是演戲，演員所做的是非常高難度的偽裝而已，他在戲中的一切喜怒哀樂、一切想法思維、一切慾望渴求，都只是高度偽裝的樣子，而不是真實的。儘管如此，一個優秀的演員可以在不斷揣摩劇情以及劇裡的角色之後，來將角色扮演得淋漓盡致，讓觀眾投入其中，以為他就是有那樣的七情六慾、那樣的慾望渴求、那樣的思維想法。

完美偽裝者論證就是要假想一個最優秀、最完美的演員。他的一切行為表現、臉部表情等，雖然都是偽裝的，但是卻跟真實的沒有兩樣。他



只是表面上被大石頭砸到了，他明明沒有處於「痛」的心理狀態，但是仍可出現「痛」之類的行為，例如皺皺眉頭、唉喲兩聲、臉部糾結等；他明明沒有那種「中秋賞月」的閒情雅致，卻能表現出他就是有那種雅興的行為。這樣的完美偽裝者，是一個沒有那些心理狀態卻有那些行為表現的人。

「完美偽裝者論證」是要指出：一個人可以表現出跟正常人一樣的行為，但是卻沒有正常人應該有的內心世界。

在提出「完美偽裝者論證」之後，帕南接著提出「超級斯巴達人論證」。他要我們想像這個情形是不是可能：有人處於「痛」的心理狀態，但是沒有顯現出「痛」之類的行為；有人心中高興極了，但是絲毫沒有表現出高興的樣子。

## 超級斯巴達人

斯巴達是古希臘時期非常著名的軍事城邦。其教育訓練是要求人人如同士兵一樣，堅忍不



拔，在最艱苦的情形下也要表現得若無其事。其實你可能看過許多人站衛兵，或者你自己就曾經站過衛兵。優秀的衛兵即使有蚊蠅爬在臉上，搔癢難耐，他還是眉頭皺也不皺，一動也不動的；他即使剛跟女朋友分手，內心難過，他不會表現出來的；他即使喜事連連，他還是一動也不動的，嘴角連個微笑也沒有；即使有小孩對他嘻笑，扮鬼臉，他還是一動也不動的。



帕南要我們進一步想像一個超級斯巴達人，他是比我們所知道的最厲害的斯巴達人、我們所知道的最優秀的衛兵，都還要厲害。他即使遭受極大的痛楚、極度的搔癢難耐、乃至於他處於任何心理狀態、任何思維，他都不會表現出一般人會表現的行為。這樣的超級斯巴達人是一個有內心狀態但沒有行為表現的人。



不論我們是接受「完美偽裝者」的可能性，還是接受「超級斯巴達人」的可能性，甚至兩種可能都接受，「痛」不會等同於任何類型的行為，「高興」也不會等同於任何類型的行為。事實上，沒有任何心理狀態會等同於任何類型的行為，也沒有任何類型的行為會等同於心理狀態。這是由於等同關係是一種必然關係。如果孔明真的就是諸葛亮，孔明就必定是諸葛亮，不會出現孔明可能不是諸葛亮的情形。同樣地，如果「痛」真的等同於某類行為，那麼痛必定等同於那類行為，不會出現「痛」可能不等同於那類行為的情形。

前面對於他心問題進行反思與檢討時，我們明顯感受到，人類大多數的行為確實是跟心理現象有關的；人類有許多行為確實是以心理狀態作為它們的原因。帕南在這裡所做的，就是回到一般人都接受的：心理狀態跟行為之間的關係應該是因果關係，而不是等同關係。這種因果關係可以受到改變，使得應該出現的行為被壓抑，或者不應該出現的行為被別的因素所產生（例如想要偽裝或模仿）。所以，邏輯行為主義將心理狀態等同於行為的學說是錯誤的。

我們可以從另一個方向來思考邏輯行為主義的不足與錯誤。對於任何一個心理狀態來說，哪些行為要列入對於它的說明？例如，在說明什麼是「痛覺」時，要將



「痛覺」等同於由哪些行為所構成的集合呢？哪些行為則不須列入？摸摸鼻子、去理髮店理個髮、狂笑等，為什麼這些行為不列入對於「痛覺」的說明？邏輯行為主義提不出原則來作決定，究竟哪些行為跟「痛」是等同的，哪些行為跟「高興」是等同的。邏輯行為主義絕對不能使用「那些由痛覺產生的行為才是用來說明痛覺的行為」來回答。別忘了，邏輯行為主義主張的是：介於心理狀態和行為之間的關係是「等同」關係，而不是「因果」關係。

其實邏輯行為主義的哲學家自始至終都沒有提出一個成功的範例，來告訴我們如何將一個心理狀態等同於一組行為或行為傾向；也沒有哲學家告訴我們究竟要如何作，才能找出這樣一個成功的範例。雖然這個批評還不足以表示邏輯行為主義就是錯誤的，在這個批評的背後有更深一層的想法值得注意：對於任何行為的說明似乎不可避免地要提到一些心理狀態。

拓夫茨大學哲學系教授丹尼特曾經指出一個一般人都願意接受的說法。「行為」有三種：本能行為、制約行為、以及意向行為。我們有很多行為跟心智的作用是有沒有關係的，像反射動作，或者生物本能的行為反應……等，就跟心理現象或心理作用無關。這些都屬於行為本能。還有一些行為則是受到環境刺激所制約的，而不是



內心的想法引起的。例如在戶外聽到上方出現聲音，就會不自主地抬頭張望，去看從頭上飛過的東西（飛機、直升機、鳥）。又例如，你穿長袖衣服的時候，習慣先套頭、再穿兩手。這個行為不是本能的，而是你自小無意養成的習慣。

此外，我們有數不清的行為是跟內心狀態有關的，這類行為稱為「意向行為」，是指透過思想或意向狀態而產生的行為。例如，我們有很多的行為是受到文化影響的，像是拜拜、祭祖、雙手合十祈禱、罵三字經、父母輕拍小孩的肩膀、男生看到漂亮女孩路過就吹口哨、情人節送紅玫瑰……等。我們有很多行為是約定的，例如手語、舉手表示贊同（或反對）等。又例如，在紙上寫一首詩、逛街、到市場買把蔥回家、報名插花班、泡一壺好茶……等。這些行為的背後都有心智的作用。

邏輯行為主義所做的是：否認意向行為，因為它否認心理狀態是行為產生的原因。所以對邏輯行為主義來



丹尼特 (D. C. Dennett)，拓夫茨大學哲學系教授，當代心與認知哲學大師之一，是哲學行為主義大師萊爾 (Gilbert Ryle) 的學生。在「意識」問題上面，反對笛卡爾式的劇場論，主張「多重草稿論」，其所著《解釋意識》(*Consciousness Explained*) 在 1990 年還進入全美非文學類暢銷排行榜，對於社會大眾與知識份子有重大的影響。



說，所謂的「行為」至多只有本能行為和制約行為兩種類型而已。然而，邏輯行為主義要如何解釋手語、舉手表示贊同、在紙上寫一首詩、逛街、到市場買把蔥回家、報名插花班、泡一壺好茶……等行為呢？這些行為很明顯都有心智的作用，似乎是很難被否定的。

## 心與腦

從以上種種反對意見來看，邏輯行為主義面臨許多理論困難，我們仍然不知道要如何說明人的心智。不過，從行為主義來論述物理論雖然失敗了，物理論並沒有因而受挫。因為物理論從腦神經科學找到了更有信服力的說法：每一種心理狀態其實都只不過是一種大腦神經狀態而已。例如「愉悅」這種心理狀態其實就是某種大腦神經狀態，「痛」則是另外一種大腦神經狀態。這個稱為「心腦同一論」的學說是 50 年代後期物理論的代表，也相當符合當代許多人的想法。可惜這個走向礙於 20 世紀初期，人類對於大腦神經系統的所知太少，也沒有足夠的理論和技術來進行大腦研究，腦科學還在萌芽的階段，所以並沒有在當時發展出關於心智的重要理論。腦科學在研究心與認知的哲學裡所佔的地位在 20 世紀中期開始，才有比較明顯的改變。隨著腦神經科學的進步，



現在有愈來愈多人相信心與腦之間有高度密切的關聯，有愈來愈多人相信心理現象就是大腦現象，有時候有人將心理現象當作是某種腦波就是一個例子。讓我們看一看心腦同一論這種說法的哲學根據。

剛剛對於心腦同一論的陳述實在是非常簡略。嚴格來說，心腦同一論主張：任何一個「類型」的心理狀態「等同」於某種類型的腦神經狀態。這裡涉及到兩個很重要的概念：「類型」與「等同」必須解說一下。

#### ①類型與個例：

首先，哲學家區別類型和個例。如果你小時候曾經被老師罰寫過字，你就懂了。例如你寫錯了「花」這個字，老師罰你在黑板上將這個字寫十遍（黑板上沒有其他任何字），你也按照老師的要求寫了十遍「花」字。哲學家說，從「類型」角度來看，你只寫了一個類型的字，那個字出現了十個個例。這是從字的形狀來分類型的。再假設你寫的前七遍「花」字是用白色粉筆寫的，後三遍「花」字是用黃色粉筆寫的。我們固然可以從字的形狀來說，這時候只有一個類型的字和十個這個類型的個例；我們也可以從顏色來說，這裡有兩個類型，白色字和黃色字，前者有七個個例，後者有三個個例。再假設你是用英文寫了三次“flower”，用中文字寫了七遍「花」



字。固然我們可以從字的形狀來說，你寫了兩個類型的字，我們也可以從字的意義來說，你寫了一個類型的字。

從這些例子要理解「類型」和「個例」的區別並不難。不過，如何對於這區別提出清楚的理論說明卻不是一件容易的事。基本上，「類型」是抽象的；「個例」是具體的，是出現於某時空位置的。

從「個例」的角度來談心腦同一，是所謂的個例物理論，例如實際在某個時間某個人感受的愉悅是「愉悅」這類型心理狀態的個例，它其實就是當時實際在那個人大腦發生的某個腦神經狀態。從「類型」的角度來說心腦同一，是所謂的類型物理論，例如「愉悅」這類型的心理狀態其實就是某種類型的腦神經狀態，「痛覺」這類型心理狀態則又等同於另一類型的腦神經狀態。哲學界所說的心腦同一論是從類型的角度來說的，心腦同一論因此也通稱為「類型物理論」。不過，反過來說，就不成立了。並不是每個類型的腦神經狀態都等同於某個類型的心理狀態，因為有些類型的腦神經狀態跟心理現象是無關的。

這個對於「類型」和「個例」的區別非常重要，不可混淆。一方面，接受個例物理論並不表示接受類型物理論（哲學家說，前者並不「蘊涵」後者），但是接受類型物理論必定接受（蘊涵）個例物理論。另一方面，當



代哲學界大致上承認個例物理論，甚至某些主張心物二元論的哲學家也同時承認個例物理論。

## ②等同：

接下來要考慮的是「等同」或者「同一」這個概念。

「x 等同於 y」跟「x 和 y 有關聯」是不一樣的，也跟「x 和 y 有因果關係」不一樣。這一點我們在之前研究行為主義時，就已經注意到了。心腦同一論主張的是每個類型的心理狀態都等同於某個類型的腦神經狀態，它並不是主張這兩種類型之間僅僅具有某種關聯而已，它也不是主張腦神經狀態在因果上會產生心理狀態。例如，當代化學主張水等同於  $H_2O$  時，並不是說水跟  $H_2O$  有某種關聯，也不是說  $H_2O$  會產生水，而是說水就是  $H_2O$ ，水和  $H_2O$  其實是一樣的「東西」。

心腦同一論在主張心理狀態與大腦狀態具有等同關係時，就是指如同水與  $H_2O$  之間的那種等同關係。不但如此，這一派的物理論哲學家主張，這種等同關係可以經由科學研究而發現，正如同水跟  $H_2O$  的等同關係是被化學所發現的一樣。心腦同一論的學說使得我們可以期待：原則上藉由腦神經科學的研究，我們應該可以發現各種類型的心理狀態與腦神經狀態之間的等同關係。也由於這個緣故，哲學與科學不再是如同一般人所認為的，



彼此是不相干的；對於心智的探索不僅需要哲學學說的建立，科學研究更有其不可或缺的重要性。

我們曾經面臨心物難題，對於心物之間的因果交互作用不知如何說明。心腦同一論的好處是：既然心智（亦即大腦）都屬於物理世界，心物因果交互作用不再是理論的困難。

儘管如此，也儘管心腦同一論看起來相當符合一些人的想法，這個學說在當代哲學卻僅僅曇花一現。為什麼呢？

在哲學界有所謂的萊布尼茲定律：如果  $x$  等同於  $y$ ，則所有  $x$  所具有的性質  $y$  都有，而且所有  $y$  具有的性質  $x$  都有。如果孔明真的就是諸葛亮，他們其實是同一個人，當然孔明所具有的一切性質諸葛亮都有，諸葛亮所擁有的一切性質孔明也都有。如果水真的就是  $H_2O$ ，它們其實是同一種東西，當然水所具有的一切性質  $H_2O$  都有，反過來說也是如此。如果愉悅這心理狀態等同於某種腦神經狀態，當然愉悅所擁有的一切性質該腦神經狀態都有，反之亦然。痛覺，乃至於所有的心理狀態都會跟它們所分別等同的腦神經狀態，滿足萊布尼茲定律。

然而事實似乎不是如此。例如，手指頭被割傷後產生了「手指頭痛」這類型的心理狀態，而且那痛覺是出現在手指頭上的痛覺。如果它等同於某種腦神經狀態，



難道該腦神經狀態也出現在手指頭上嗎？這是說不通的。除非我們認為手指頭痛的時候，那個痛覺不是出現在手指頭上，而是出現在腦裡，將痛覺等同於腦神經狀態才不會遇到這個困難。可是，手指頭被割傷後的痛覺真的出現在腦裡，而不是在手指頭上嗎？痛覺是具有空間位置的，這一點似乎不太有人質疑，痛覺可以出現在大腦以外的身體部分，幾乎沒有人會反對。牙痛的部位在蛀牙的地方，胃痛的部位在胃部。兩者都不是出現在大腦。所以依據萊布尼茲定律，痛覺不應該等同於大腦神經狀態。

除了痛覺、癢覺之類的感質具有空間性，因此無法等同於大腦神經狀態之外，在第二章我們已經注意到，心物難題之所以出現的一個重要理由是：心理世界具有一些重要的特徵是物理世界不會有的，例如私密性、主體性等。所以，主張心跟腦是同一的，仍然會違反萊布尼茲定律，因為大腦不會具備私密性和主體性等性質。我就不再多說。

除了考慮萊布尼茲定律之外，哲學家指出：「等同」是具有遞移關係的；換句話說，如果  $x$  等於  $y$ ， $y$  等於  $z$ ，則  $x$  等於  $z$ 。假設狗的痛覺跟人的痛覺是屬於同一類型的。如果人的痛覺等同於人的某類腦神經狀態，狗的痛覺等同於狗的某類腦神經狀態，難道人的那類型腦神經



狀態等同於狗的那類型腦神經狀態嗎？如果海豚也會有痛覺，難道人、狗和海豚都具有相同的那類型的腦神經狀態嗎？事實上這是錯誤的。心腦同一論顯然違背了腦神經科學對於人、狗和海豚，乃至於其他動物的大腦研究。經由科幻電影的刺激，許多人常常認為如果有外星人，他們也是有心智的，也會有痛覺和各種心理狀態的。但是除非我們所說的外星人跟地球人在大腦方面沒有太大差別，否則跟地球人大腦結構差異懸殊的外星人會有心智，會有痛覺、會想征服地球……，就變得難以想像了。

或許一開始我們就不該假設狗的痛覺跟人的痛覺是相同類型的。狗跟人的行為反應差異太大了。同樣是被石頭砸到腳，人會抱著腳一面跳一面哇哇叫，狗當然做不到。而狗的哀叫聲和竄逃的動作也不是一般人的反應方式。所以我們憑什麼假設狗的痛覺跟人的痛覺屬於相同類型呢？

如果你有這種想法，反而更能指出心腦同一論的弱點。因為你顯然注意到了即使不同類型的生物，即使反應差異極大，即使大腦結構差異極大，還是可以擁有痛覺的。即使是不同的生物也可能擁有跟人類一樣相同的心智。動物是不是有心智，我們會在第八章作一些探討。然而是不是必須承認只有人類之類的大腦神經系統才能



擁有心智呢？有沒有可能擁有不同大腦結構，甚至是由不同物質構造出來的大腦的別種物體，例如外星人，也能跟我們人類一樣擁有相同的心智呢？甚至人造的機器人，完全使用金屬、半導體、生化科技產品等製造出來的、用電腦無線操控的機器人，有沒有可能也和我們人類一樣擁有相同的心智呢？

當代電腦的發達，使得愈來愈多的人相信人心如同軟體，人腦如同硬體。這當中的一個想法就是：只要硬體的使用能夠作到執行程式無礙，即使所採用的硬體種類不同，不會影響程式本身。「人工智能」就是一門企圖從程式的角度來理解人類智能的學問；它的目的簡單說，就是造出一臺和人類一樣會思考的機器人。也正因為如此，只要所找到的程式確實是思考的程式就行了，至於用什麼物理材質來執行程式，是不是要將機器人造得跟真正人類外型上一模一樣，倒不是重點。然而如果心腦同一論是對的，諷刺的是，人工智能不僅是不必要的，人工智能學者甚至走了錯誤的研究方向。我們該何去何從呢？在我們進一步瞭解人工智能的發展時，讓我們繼續看其他的哲學家如何看待大腦神經系統在瞭解心智時所扮演的角色。



## 大家都錯了！

在哲學界有一個非常奇特的立場：大家都錯了！其實人類沒有心智！我們一般所說的相信、希望、判斷、思考，我們一般人所經歷的七情六慾、喜怒哀樂，都是虛幻的。乍聽之下，這個說法著實令人錯愕。強邏輯行為主義固然是這種主張，加州大學聖地牙哥校區哲學系教授丘曲冷的取消唯物論更是有名。他主張：在發展理論以瞭解人類的過程中，我們需要的不是心理學，而是腦神經科學。



丘曲冷 (Paul M. Churchland) 當代心與認知哲學的主要人物。當代取消唯物論能夠自成一家之言，他的貢獻功不可沒。撰有《科學實在論與心智的彈性》、《神經計算觀》等重要著作。

在說明他心問題時，我們曾經提到老百姓心理學或者常識心理學。常識心理學主要是針對認知主體所相信的事、所慾求或希望的事，以及這兩類心理狀態所導致的行為，來建立心理因果定律。所以又稱為信念—慾望心理學。正如之前所說的，常識心理學承認人有心智，人會思考、判斷、慾求、想像……等，這些心理狀態會產生行為，而且它



們跟行為之間的聯結具有規律性。常識心理學描述了這些「心理—行為」因果定律或通則。這些因果定律或通則是一般人透過對於「心理—行為」因果關係的觀察，經由歸納，來找出心理現象與個體行為之間規律的通則。這些定律或通則雖然沒有寫進教科書，雖然沒有系統，也幾乎不曾被清楚地說明過，甚至也很少明白出現在一般人心中、被一般大眾所察覺到，這些定律或通則卻是一般人日常生活所奉行不渝的。我們會用這些定律或通則來解釋、預測別人的行為，而且大致上相當成功。

然而丘曲冷認為常識心理學根本就是個錯誤，沒有信念、沒有慾望，一切我們一般人自以為擁有的內心世界，都是虛幻的。他認為常識心理學是一個理論，而且是一個很糟糕的理論，因此應該拋棄這個理論以及它所承認的心理世界。取而代之的，應該是腦神經科學。我們現在雖然使用心理語言，像是「高興」、「痛苦」、「思考」、「判斷」等，那是迫不得已的，因為我們還沒有發展出適當的詞彙。我們現在誤以為人類有心智，誤以為我們的心理世界是真實的，但是遲早有一天，當腦神經科學的研究全部完成之後，我們就可以不必再使用心理語言，我們也不會再錯誤地將心理世界當作是真實的。這情形就跟當年物理學家一樣，原本承認以太的存在，並藉著以太理論來說明一些物理現象（以太是光傳播的



媒介)，但是後來科學家放棄了這種想法，也不再承認有所謂的以太。

丘曲冷憑什麼要否認心理世界呢？他為什麼認為常識心理學是錯誤的呢？他提出了以下的理由：(1)常識心理學數千年來一直是停滯而沒有進步的；(2)常識心理學的解釋力非常貧乏；(3)常識心理學與科學史上曾經出現過的煉金術、燃素論、和以太論等理論，具有相當程度的類比性；以及(4)就對於行為原因的解釋來說，一個不按照常識心理學而建構出來的神經科學，其先驗的成功機率，比起想要按照常識心理學來建構的神經科學要高得多，因為概念上有太多種不按照常識心理學來建構神經科學的方式。讓我逐一說明。

丘曲冷的第一個理由相當清楚。幾千年以來，人類不分東西方，確實一直是使用七情六慾、信念慾望等來解釋和預測人類的行為。就這點來說，常識心理學確實是沒有進步的。他的第二個理由是常識心理學的解釋力非常貧乏。他舉了一些常識心理學無法提供解釋的例子：常識心理學不能解釋心理疾病、不能解釋我們三度空間的視覺、不能解釋創造想像力的性質、不能解釋大腦受傷與心智運作失常之間的關連……等。常識心理學的解釋力似乎非常薄弱。

丘曲冷的第三個理由我們剛剛已經看到了。在科學



史上確實有許多理論曾經獲得多數科學家的支持，並且對於研究課題提供了不錯的解釋和預測。大多數當時的科學家也都接受這些理論所承認的事物或現象是真實的。但是隨著科學的進展，有許多理論後來都遭到淘汰了，也不再有任何科學家接受原本承認的事物或現象。燃素論和以太論就是最好的例子。丘曲冷認為常識心理學就跟這些理論一樣，而認定常識心理學只有走上被淘汰一途。



先驗 (A Priori) 一詞簡單講，說一件事情是先驗的，就是說對於這件事情的瞭解完全不需要透過觀察、不需要做科學研究。通常數學知識和邏輯知識都是先驗的，另外，像「沒有任何平面的形狀既是圓形又是方形」也是先驗的。

丘曲冷的最後一個理由比較特殊，他認為在觀念上有太多種不使用常識心理學來建構神經科學的方式。為了方便起見，讓我們假設有甲、乙兩種建構神經科學的方式，是不使用常識心理學的。所以連同常識心理學，一共有三種建構神經科學的方式。觀念上來說，用常識心理學成功建構出神經科學的機率是三分之一，用不是常識心理學來作，成功的機率是三分之二。所以不按照常識心理學來建構神經科學，其成功的機率大於只按照常識心理學來建構的作法。這是丘曲冷所謂的「先驗的成功機率」。

根據以上這些理由，丘曲冷於是主張我們應該廢棄



常識心理學，連帶將它所承認的喜怒哀樂七情六慾、和思維判斷等種種心理現象，一併加以否認。所以，大家都錯了！我們沒有心智！

我們能接受丘曲冷的主張嗎？檢討他的理由，說實在的，並不是那麼令人信服。例如，常識心理學停滯不進步的說法並不足以支持它是不好的一門「學」。其實，一般人使用信念和慾望作解釋預測的作法，為什麼是一門學、一個理論，本身就是有疑問的。更何況就算一個理論是沒有進步的，也很難說這個理論就是錯誤的。只有在一個理論不但停滯不前，而且沒有足夠的解釋和預測力，我們才會傾向於認為它是一個不好的理論。但是丘曲冷提出的常識心理學解釋力薄弱的例子，實在讓人不滿意。對於心理疾病、三度空間的視覺、想像力的性質、大腦受傷與心智運作失常之間的關連等現象，常識心理學確實無法提出解釋，但那是因為常識心理學本來就不是針對這些現象而提出解釋的理論。我們一般人使用信念慾望等，是用來解釋或預測人的行為的，僅此而已。你會因為物理學沒有辦法解釋「演化」或者人類的大腦結構，就說物理學是不好的理論嗎？至於第四個理由也是相當薄弱的。畢竟沒有人知道要用什麼方式來建立真正的神經科學，也沒有人要求在建立神經科學的過程中，必須要使用「信念」、「慾望」等常識心理學所使



用的概念。神經科學的建立，或者認知神經科學的建立，是不是要搭配常識心理學呢？我們需要更多的思考，才能清楚其中的關聯。

取消唯物論這種極端的立場是失敗的。它的失敗不是由於它是極端的，而是由於它沒有提出令人信服的好理由。

我們大致上看過了心腦同一論以及取消唯物論。雖然我們有相當的理由不接受這兩個理論，我們並沒有排除腦神經科學對於心智的研究所可能提出的貢獻。下一章我會從「功能」這個全新的角度來繼續思考心智的問題，我們會再繼續思考腦神經科學的研究可以扮演什麼角色。

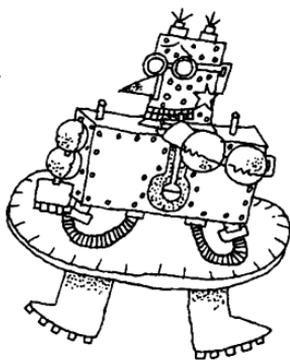


## 5 會思考的機器人

機器人不  
可以思考  
嗎？

如果有一  
天，我像駭客  
任務的機器人  
一樣厲害，

人類就會  
和我做朋  
友了嗎？





## 瘋子的痛與外星人的痛

心腦同一論和取消唯物論都指出了大腦神經系統對於瞭解心理世界的重要性。取消唯物論由於所提的理由不夠具有信服力，並沒有得到大多數人的支持；心腦同一論雖然面臨極嚴重的挑戰，但它之所以那麼快就被哲學家放棄，主要還是由於當時新出的理論更吸引人。

我們固然認為心理現象跟人的大腦神經系統有密不可分的關聯，我們也常常覺得即使是不同的生物也可能擁有跟人類一樣的心智。由於當代電腦的發達，我們甚至開始覺得或許人的心智如同軟體，人的大腦如同硬體。IBM 的深藍電腦（程式）能夠跟世界西洋棋王對弈的事情讓人印象深刻。或許將來我們能夠寫出跟人類思考相似的電腦程式，將它安裝在人造的、外型跟真人難以區別、外表摸起來跟真人很像的假人。然後只要執行這套「思考程式」，就可以操控假人的手腳、它的臉部「肌肉」的動作、乃至於發出跟真人一樣的聲音。尤其這假人的行為由於是由那套「思考程式」控制的，所以它的行為跟真人的行為應該沒有什麼不同。科幻小說和電影中，對於具備思考能力的機器人的描述，栩栩如生，似乎那種機器人的出現只是早晚的事，就只等人類科技如何突



破而已。比電影《星艦迷航》裡的機器人百科，更像人類的機器人應該是指日可待的。

然而大腦神經系統呢？我們不也是認為心理現象跟大腦神經系統有密不可分的關係嗎？如果只要設計出那個程式就可以有思想了，電腦的硬體似乎沒有扮演太重要的角色。要知道那「思考程式」是什麼，並不需要研究電腦硬體，僅僅研究電腦硬體也不會告訴我們那「思考程式」是什麼。既然如此，在我們試圖瞭解人類的心智時，似乎並不需要對於大腦進行研究。腦神經科學的研究對於心智的瞭解似乎是不相干的。

這個困境由於功能論的提出而得到了解決：每一種類型的心理狀態都等同於某種類型的功能狀態。哲學家認為「心」這個介於環境刺激和行為反應之間的黑盒子，要用功能角色（亦即因果角色）來理解。各種不同的環境刺激、行為反應以及心理狀態，織成了一個因果的網絡；每個心理狀態都跟環境刺激、行為和其他的心理狀

態之間具有因果關係。例如「痛」雖然不等於某組行為，但是「痛」指的是跟某組環境刺激、某組行為以及其他心理狀態有因果關係的一種心理狀態。由於這個緣故，這派學說又稱為因果角色功能論。



功能論有許多派別，本書只介紹比較知名的因果角色功能論以及算機功能論。



因果角色功能論的說法其實還可以更嚴謹一些。嚴格來說，每類心理狀態不是等同於某類因果角色，而是等同於扮演那類因果角色的那類事物，而且有可能不同類型的事物可以扮演相同的因果角色。這個定義有些晦澀，讓我舉一個例子做類比，或許會比較容易理解。

每一齣戲都有劇本，劇本描寫了戲中每個角色會做的事，至於每個角色由誰來演出的，則不是劇本會過問的。我們可以這樣說：劇本的描述決定了劇中角色的行為，演出者則負責在劇中將這些行為演出來。如果劇本修改了，有可能是增加了新的角色，有可能是取消了某些角色，也有可能是要求原有的角色在劇中做一些不同的行為。此外，在拍戲的時候，如果某個角色的演出者不稱職，當然可以將他撤換，另外找別人來演出相同的角色，演出劇本所要求做出的行為。我們對於一齣戲區別了「角色」以及「角色扮演者」。

因果角色功能論也做了類似的區別：「痛覺」是由一組因果角色定義的，「扮演痛覺這個角色的」是某種事物，通常是指人的大腦狀態；不同的因果角色是由不同的大腦狀態扮演的。以「痛」作例子。假設「痛」是由 R 這個因果角色定義的。對人類來說，就是某類大腦神經狀態扮演了 R 這個因果角色，痛等於人類那一種腦神經狀態。現在，類比來說，在一個劇本裡，相同的角色可以



由不同的人來演出，因此相同的 R 這個因果角色，也可以由不同的腦神經狀態來扮演。這個說法使得我們在理論上可以承認：如果有外星人，如果外星人有跟人類一樣的痛覺，那麼即使外星人的大腦構成的物質以及大腦的結構都跟人腦不同也沒有關係，因為痛覺是由相同的因果角色 R 定義的，外星人的「痛」等同於他的那種腦神經狀態，地球人的「痛」等於地球人的某種腦神經狀態。心腦同一論的困境似乎解除了。

現在的困難是：要如何找出心理狀態的因果角色？我們曾經提到的常識心理學，在這裡就派上了用場。大致來說，一般人都承認被球狠狠打到腳時會讓腳痛，這腳痛會引起揉一揉腳的動作、皺皺眉頭、唉呦兩聲等行為，乃至於心理埋怨著：「真倒楣！」所謂的「痛」就是由這些環境刺激、行為反應和其他心理狀態之間的因果網絡所定義的。而且定義「痛」的因果角色是由人類的大腦狀態來扮演的，關於人類腦神經科學的研究也因此而有其理論上的必要。又例如一般人承認你如果跟朋友約在某時某地見面時，除非你有意欺騙，或者你有事耽擱、或者你忘記了，否則你是會準時出現在相約的地方的。還有很多很多，都是我們對於環境刺激、心理狀態和行為反應之間因果關係的觀察與瞭解。比起邏輯行為主義、心腦同一論、以及取消唯物論，因果角色功能論



對於心智的說明似乎更符合一般人的想法。

因果角色功能論還有一個好處：它在解除心腦同一論的困境同時，還能承認大腦神經系統的重要性。它還允許我們發揮無盡的想像空間。正如同在一齣戲裡，同樣的角色可以由不同的人來演出，同樣的因果角色也可以由不同的大腦結構來扮演。即使有外星人存在，即使他們的大腦是由特殊的材質構成的（例如由純銀構成），跟地球人截然不同，只要那種材質所造成的外星人大腦能扮演相同的因果角色，外星人就會有心智。如果有一天科學發達了，我們找到了各種心理狀態的因果角色，即使我們使用塑膠來作機器人的大腦，只要作出來的能扮演相同的因果角色，這種塑膠大腦機器人也會有心智，會跟人類一樣擁有喜怒哀樂七情六慾，會思考、判斷、推論、想像。

儘管如此，在我看來因果角色功能論似乎隱藏了太多的預設。我對於這個理論有許多質疑。我們真的可以承認外星人跟人類一樣有心智這種可能性嗎？因果角色功能論憑什麼認為外星人的大腦能夠扮演跟人類大腦相同的因果角色呢？別忘了對於因果角色的瞭解必須依賴常識心理學。但那是我們人類的常識心理學。就像剛剛舉的例子：我們大致上願意承認，一般人在被球狠狠打到腳時，會造成腳痛，而且這腳痛會引起揉一揉腳的動



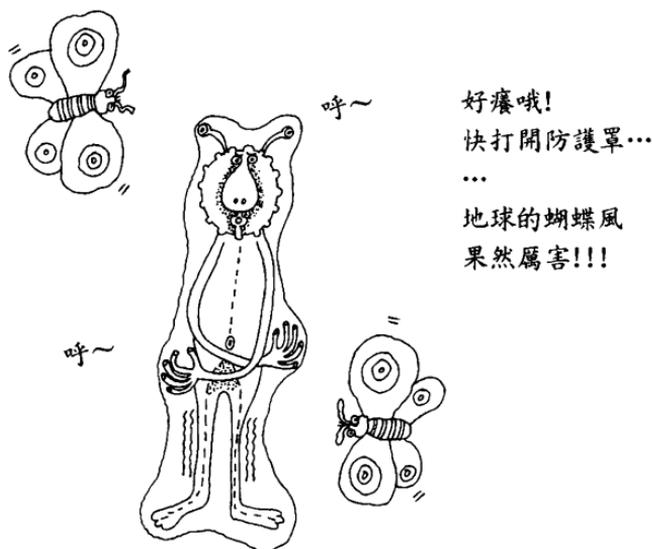
作、皺皺眉頭、唉喲兩聲……等行為，乃至於還會心裡不太高興、有點抱怨。在人類的常識心理學裡，所謂的「痛」就是由這些環境刺激、行為反應和其他心理狀態之間的因果網絡所定義的。然而我們憑什麼認為外星人被球狠狠打到腳（假設他們有腳），造成腳痛（假設他們會覺得腳痛），他們會出現跟人類相似的行為反應？他們會揉一揉腳、皺一皺眉頭、唉喲兩聲……？那是我們人類的常識心理學，不是外星人的。我們憑什麼認為我們人類的常識心理學能夠適用於所有的外星人？難道我們如此傲慢？更何況即使真的有外星人，即使外星人真的也有心智，即使外星人也會覺得痛，跟他們的痛覺有關的環境刺激、行為反應和其他心理狀態之間的因果網絡，跟人類的未必相同。

或許你想到可以對於因果角色功能論作一些修改。我們人類的痛和任何心理狀態是靠我們人類的常識心理學所發現的因果網絡來定義的，外星人的痛和任何心理狀態是靠他們的常識心理學所發現的因果網絡來定義的；如果有不同種類的各種外星人，那麼不同種類的外星人的「痛」都是由各自的常識心理學所發現的因果網絡來定義的。這樣一來，因果角色功能論的確免除了上一段所提出的理論困難。

不過我們未必願意付出這種修改所要花的代價：即



使真的有外星人，即使外星人真的也有心智，即使外星人也會覺得痛，當外星人出現在我們人類面前時，我們



永遠不可能知道究竟外星人是處於痛覺、快樂、或者任何心理狀態。因為按照這個修改的說法，外星人的心理狀態所扮演的因果角色，是由他們的常識心理學決定的，但我們人類如何學到他們的常識心理學呢？如果我們不知道他們的常識心理學，我們要如何知道某個外星人正在覺得痛、正在感到高興呢？

事實上我認為我們遇到的困難比這更嚴重。如果我們用因果角色來定義「痛」，因果角色用常識心理學來決



定，而不同的族類（外星人）有不同的常識心理學，我們就會推導出：不同族類的痛覺是由不同的因果角色來定義的。這一點卻違背了因果角色功能論的基本主張。當外星人覺得痛的時候，不僅他的生理物理材料和構造跟人類不一樣，他的大腦狀態跟我們的不一樣，就連他的「痛」的因果角色也跟人類「痛」的因果角色不一樣。那麼，外星人的痛跟人類的痛還有什麼地方是一樣的？外星人的歡欣跟地球人的歡欣還有什麼地方是一樣的？外星人的任何一個心理狀態跟人類相對的心理狀態還有什麼地方是一樣的？如果因果角色功能論仍然要承認外星人也可以有痛覺和各種心理狀態，它只好承認這樣的外星人一定是在外型和內心世界都跟人類非常相像的外星人。

同樣的疑難其實不必等到考慮外星人才會出現。在我們人類當中，有一些人被視為「瘋子」。不論他們是為了什麼原因發瘋的，他們的行為總是極端異於常人，他們的思考也極端異於常人，他們的喜怒哀樂也是極端異於常人。簡單說，即使人類的常識心理學對於人類的瘋子也是派不上用場的。這就是為什麼我們無法預測瘋子的行為，也無法解釋瘋子行為的原因。然而即使如此，瘋子也是有心智的，因為瘋子也有喜怒哀樂、七情六欲。因果角色功能論要如何說明瘋子的喜怒哀樂七情六慾



呢？瘋子的痛要如何理解呢？外星瘋子呢？他們的心智要如何瞭解？他們的思維要如何瞭解？他們的喜怒哀樂要如何瞭解？

對呀！對呀！  
瘋子也是有心智的呀！  
哎唷！好癢哦！



經歷了行為主義和心腦同一論，好不容易因果角色功能論帶來了一道曙光，外星人的痛和瘋子的痛又把我們拉到了困惑無解的境地。我們只好再繼續思考其他的出路了。

## 圖林算機

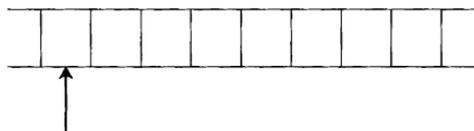
功能論的型態很多，除了剛剛提到的因果角色功能論之外，跟當代認知科學和人工智能的發展有關的，是哈佛大學哲學家帕南從計算理論提出來的算機功能論。從這個角度思考心智問題的哲學家，所試圖回答的是：



有可能造出一臺會思考、具備心智能力的機器人嗎？關於具備思考能力機器人的想法，並沒有僅僅停留在科幻小說和電影裡。早在 60 年代還沒有出現這類科幻小說和電影的時候，哲學家就已經從計算理論出發，從純粹程式的角度來瞭解心智。只要硬體的使用能夠作到執行程式無礙，即使所採用的硬體種類不同，也不會影響程式本身的設計。現在電腦的發達與普及更讓人覺得這想法是很自然的。

算機功能論嚴格來說，並不是物理論，也不是唯心論，但卻是一種可以跟物理論相容的主張。這個哲學主張為人工智能的發展提供了完整的概念架構與理論基礎，並且在電腦發明之後得到了更廣泛的支持。

要瞭解算機功能論，首先就要知道什麼是算機。這裡所說的「算機」不是硬體的概念，不是指我們一般所說的，那些有金屬外殼、內裡有各種零件的機器。「算機」是純粹抽象的概念，指的是所謂的「圖林算機」。在哲學和認知科學裡，圖林算機是一種抽象的程式，通常是用所謂的「算機表」來表示；但也有使用所謂的「有限四項式序列」以及「流程圖」來表示的。圖林算機的執行通常是採用一個假想其左右兩邊可無限拉長的細長空白帶，以及一個負責讀取資料與輸出資料的「讀寫頭」（以↑表示）：



對於算機的瞭解必須包括四件事：

- (a) 輸入資料與輸出資料的類型
- (b) 算機的起始狀態
- (c) 算機的結束狀態
- (d) 算機所處的各种狀態（即該算機的讀寫頭遇到格子裡的資料時，所應做出的動作）

通常算機使用「1」和「0」作為基本的輸出入資料的類型。連續  $n$  個「1」就構成一個輸入（或輸出）的字串（ $n$  為大於或等於 1 的整數）。（理論上，輸入和輸出的資料類型並不一定是以 0 和 1 來編碼。）其次，算機的起始狀態和結束狀態都是指讀寫頭指著這些連續  $n$  個 1 的最左邊的 1（也可以是最右邊的 1）。而讀寫頭除了讀或寫格子的資料之外，它還可以做出每次往左一格或每次往右一格的動作。我們平常所謂的電腦，乃至於自動販賣機，其實都不脫離圖林算機的理论架構。

計算理論是一門專技的領域，不是這裡一兩頁的篇幅就能介紹清楚的，所以我試著用例子來說明算機功能論的要旨。讓我們考慮一臺功能相當簡易的自動販賣機，



它只接受 5 元硬幣和 10 元硬幣，它只販售每罐 10 元的飲料。底下是為這臺販賣機所設計的圖林算機，用算機表來表示（這個例子源自於紐約大學哲學系教授布拉克的可樂販賣機）：

	輸入 5 元	輸入 10 元
狀態一	(a) 1. 不做動作 2. 進入狀態二	(b) 1. 放出飲料一罐 2. 停留在狀態一
狀態二	(c) 1. 放出飲料一罐 2. 回到狀態一	(d) 1. 放出飲料一罐 2. 退回 5 元 3. 回到狀態一

這臺販賣機的起始狀態就是「狀態一」。假設張三先投入了一個 5 元硬幣。由於每罐飲料都是 10 元，所以這機器沒有動作，但是因為張三已經投入了 5 元硬幣，所以依照 (a) 格的指令，這臺機器已經進入了狀態二。張三接著有兩種作法，一是再投入一個 5 元硬幣（此時要考慮 (c) 格），一是另外投入一個 10 元硬幣（此時要考慮 (d) 格）。如果張三接著投入的是一個 5 元硬幣，依照 (c) 格的指令，這機器放出一罐飲料並回到狀態一；交易結束。但是如果張三接著投入的是 10 元硬幣，依照 (d) 格的指令，這機器會放出一罐飲料，退給張三 5 元，然後回到狀態一；交易結束。



如果王五也來用這臺販賣機（狀態一），但他一開始投入的就是 10 元硬幣。這時候依照 (b) 格的指令，他買到了一罐 10 元的飲料，仍然停留在狀態一；交易結束。

這臺販賣機用一些簡單的技術和材料就可以實際做出來。但是顯然關鍵不在於硬體，而在於如何設計。其他更複雜的販賣機的製造，在抽象層面上所使用的原理都是一樣的；只有在實際製作時，才會考慮所使用的材料、所接受的輸入類型、以及輸出類型。例如利用相同的原理，我們可以造出販賣飲料、速食麵、餅乾……等的機器，而不限於販賣飲料；我們的機器也可以更方便一點，來接受硬幣、紙鈔、磁卡、鍵盤上的按鍵……等。不但如此，這個原理不一定要拿來製造販賣機，也可以拿來製造點唱機、CD 試聽機，乃至於更複雜的電腦象棋、PDA……等。

所有這些抽象意義下的算機（也就是程式）是所謂的符號操作系統。算機（程式）的寫法是邏輯的、語法的，算機所操作的是符號，而且嚴格來說，在操作符號時，並不是依據符號的意義，而是純粹就符號的物理性質來執行程式的。以剛剛介紹的算機理論來說，那個空白帶所處理的輸入，嚴格來說，不是「1」和「0」兩個自然數，而是刻畫在帶子上的那兩種形狀。寫算機（程式）時，改用別的形狀，甚至別的方式照樣能執行相同



的功能。例如，我們可以用「燈泡亮」來取代「1」，用「燈泡不亮」來取代「0」。

再以那臺販賣機來說，它的輸入，表面來說只有 5 元硬幣和 10 元硬幣兩種，但其實它處理的是不同大小的圓形硬物。至於那兩種圓形硬物是不是真的錢幣，機器並不在乎（如遊樂場中的代幣）。算機的運作就是根據物體的物理性質來執行的。哲學家說，這些物體的物理性質具現（具體實現）了算機程式。請讀者留意的是：這種「具現」的關係不是等同關係，也不是因果關係。硬體所處的狀態當然不是程式所處的計算狀態，硬體所處的狀態也不會產生程式的計算狀態。從前的電腦用的是體型巨大的真空管，現在的電腦短小輕盈，用的是矽晶片，將來或許還會使用更新穎、更精緻的不同材質來做電腦。這些材料的使用不僅是體積的差異，更是使用了不同的物理學原理。由於程式具有可被不同硬體執行的性質，哲學家將這種性質稱為多重可具現性。當然，既然同樣的算機程式允許使用不同的材質，而不同的材質使用了不同的物理學原理，我們是不可能從硬體的物理結構和物理性質來找出它們在執行的是什麼程式的。



圖林算機 (Turing Machine) 「算機」這個概念是英國數學家圖林 (A. M. Turing, 1912-1954) 提出的。圖林畢業於倫敦國王學院，主攻數理邏輯。在二次大戰期間被英國政府徵召，擔任破解德軍密碼的工作。在圖林的顛峰時期，他被指控同性戀，於 1954 年自殺去世。他所提出的「算機」概念提供了 1940 年代電子數位計算器（也就是俗稱的電腦）的理論基礎。圖林認為人類的行為可以用「算機」概念來理解，這是後來心與認知哲學界發展出算機功能論的濫觴。



## 人與機器

現在讓我們回頭來看看我們原先所關心的問題。在 60 年代的時候，帕南主張：圖林算機的概念架構提供了我們如何理解人類心智的一個最好的理論。人的心智系統其實就是一個圖林算機！人的每一個心理狀態就是一個計算狀態；人的心智歷程就是算機的計算歷程，並且以所謂的心理表徵作為機器操作的對象（見第六章）。以上面販賣機的例子來說，它的算機表界定了它的計算歷程，而 5 元和 10 元那兩種大小的圓形硬物，就是該販賣機在運作的時候所處理的表徵。

哲學界將這種用「算機」概念來理解人類心智的主張稱為算機功能論（又稱為計算心理學派或者古典學



派)。根據這套學說，心智就是一套符號操作的系統而已。這派學說也開啟了後來對於人工智能的研究。如果算機功能論是對的，則或許真的有一天我們能造出一臺會思考、具備心智能力的機器人！認知科學家的工作就是去找出（寫出）人類心智的圖林算機；工程師的工作就是設法將這些想法用最新的技術與材料實際將它建造出來。

話雖如此，事情沒有這麼簡單。我們一直深深關切著人心與人腦之間的關係。很多人認為人的心智其實就是大腦的活動。有很多人認為人腦跟人的心智有密切的關聯，有許多人認為對於人的心智進行研究時，對於人腦的研究是不可或缺的。某些部位的腦傷可能引起認知的障礙，不是嗎？那麼，人的心智與人腦之間又有什麼關係呢？或許你認為這不是什麼大不了的問題。既然將心智類比為電腦的軟體，將大腦類比為電腦的硬體，硬體故障當然會造成軟體無法執行。

真正的問題比這深多了。在主張算機功能論的哲學家來看，跟電腦對照的說法已經不只是比喻而已，事實上人的心智就是軟體，人的大腦就是執行這軟體的硬體。這個說法所遇到的狀況是：如果人的心智就是圖林算機這種抽象的程式，如果心理狀態等於計算狀態，要研究人的心智，所需要的是研究那抽象的程式，不是那執行



程式的人腦。正如同純粹只從物理學的角度研究你家電腦的硬體部分，沒有辦法讓你知道任何關於它所執行的軟體，純粹從腦神經科學的角度來研究人腦這個硬體，應該也沒有辦法讓你知道任何它所執行的軟體（構成人的心智的圖林算機）。那麼，人腦在算機功能論中的理論地位是什麼呢？

根據算機功能論的主張，既然人的心智就是那抽象的圖林算機程式，理論上這個程式也允許由不同的硬體來執行。人腦具體實現了那個「心」的程式。人腦跟人心的關係不是等同的關係，不是因果關係，而是實現的關係。跟因果角色功能論一樣，算機功能論這個說法允許我們使用不同物理性質的材料，依據不同的物理學原理，來作為執行程式的硬體。前面提到的販賣機就是一個很好的例子。

算機功能論將人的心智等同於一種圖林算機，人的心智也具備這種多重可具現性。這說法聽起來蠻好的。當人工智能的研究有了具體成果之後，如果要製造一臺會思考的機器人，我們不一定要給它一個跟人腦相同的「大腦」，我們可以用一堆矽晶片、金屬、橡膠……等，或者未來世界的材料，設法建造出能執行人心智程式的硬體。不僅如此，科幻小說和電影常常有關於外星人的情節。不論故事情節是邪惡外星人入侵地球，還是善心



外星人跟人類和平共處，電影和小說都設想那些外星人具有相當的心智能力，但是同時它們的身體和大腦都跟地球人截然不同。多重可具現性的說法，竟然也使得科幻小說和科幻電影有了學理的基礎。

讓我將**算機功能論**的基本主張摘要如下：

- (1) 心智是圖林算機；或者說心智是一種符號操作系統；
- (2) 心智歷程是一種以心理表徵為單位的計算歷程；
- (3) 心智具有多重可具現性，在人類的情形裡，心智是被人的大腦具現的，但不一定非要由人腦才能加以具現。不同的材質、不同的大腦結構、乃至於外星人，只要能具體實現「心智」的算機程式，都可以擁有心智。

在我繼續說明有關這個學說的評論之前，讓我們回到之前我在解說行為主義的時候，所提到的「圖林測試」。我們到目前為止對於算機功能論的說明都是在回答心智是什麼的問題。圖林測試則不是在回答這個問題。它已經假設算機功能論是正確的哲學立場，接受了程式的執行就是心智運作的主張。但是，畢竟不是所有程式都是真正寫出心智的程式，所以必須要有一套方法來決定，程式設計者所撰寫的程式是不是真的就是心智的程式。這套方法就是圖林測試。圖林測試的作法大致如此：



將執行程式的電腦放在某個房間，不讓評審知道所面對的是真人還是電腦。然後由評審來提問題。由於科技還不夠發達的緣故，為了避免因為電腦發出的聲音而使得評審知道所面對的是電腦，所以我們可以改用文字輸入與輸出的方式來進行測試。如果評審無法辨別所給的回答究竟是真人的答案，還是電腦的答案，就表示這個程式通過了圖林測試；如果評審能夠察覺那不是真人的答案，就不算通過測試。

基本上圖林測試是採取行為主義的研究方法。我們曾經看過心理因果原則，這原則說，心理狀態跟行為之間具有因果關係。常識心理學也提供了我們在正常情形下，什麼行為跟什麼心理狀態具有定律式的（或者通則式的）關係；功能論就是從這個角度來定義心理狀態的。既然如此，藉著圖林測試似乎可以從行為來反推行為者的內心想法。到目前為止，圖林測試在人工智能界仍然有很重要的地位。

## 不懂中文的洋人

算機功能論對於當代心與認知研究的影響可說非常



巨大。不過在 80 年代史坦佛大學哲學系教授瑟爾提出了相當有名的批評，稱為中文房間論證。他要我們依據圖林算機的基本概念來想像一個房間：

房間裡有一位只懂英文、不懂中文的美國人。這位美國人手裡有一份手冊，是已經設計好的程式。手冊是用英文寫的，所以他看得懂裡面的指令。從房間外面會送進卡片，上面寫著中文。這位美國人的工作是依據他手中的手冊來做出適當的動作，包括向房間外面遞出上面有寫著中文的卡片。

這位美國人其實就是充當圖林算機的讀寫頭。每張卡片相當於圖林算機的輸入，上面寫的中文字類比為空白表格裡的符號。瑟爾說，由於確實有人懂中文，因此，按照算機功能論的主張，原則上我們應該能寫出一個算機程式，來說明人類是如何理解用中文寫的小說。假設我們真的將這臺算機寫了出來。按照程式指令我們將用中文寫的卡片作為輸入，交給房間裡的那個美國人，按照程式的設計，房間裡的美國人會將適當的輸出（也是用中文寫的卡片）傳到房間外。但是瑟爾要我們想一想：在他的故事裡，這種理解中文的心智能力究竟是出現在哪個地方？絕對不會是那位美國人，因為早就已經規定



好，我們找的人是只懂英文、不懂中文的。我們找的是一個即使不懂中文也能執行程式的人。這個人只是在依據手冊的說明，針對他所看到的卡片上符號的形狀，來執行程式的指令而已。就如同販賣機不知道它所處理的是金錢和飲料一樣，這個美國人並不懂得所操作符號的意義。不但如此，這個美國人當然也不會因為讀得懂英文的手冊就讀懂了一本中文小說。那麼，難道我們要說那間鋼筋水泥蓋的房間懂中文嗎？好像也不對。或許是寫手冊的那個人吧？但是那個人自始至終都沒有出現在故事裡，而且跟我們的問題一點也不相干。那麼，到底我們要不要說「懂中文」這個現象，的確有出現在這個故事裡？

如果我們否認在這故事裡有出現「懂中文」這個現象，那麼我們就等於是否認了設計圖林算機來解釋「懂中文」這個心智現象的可能性。中文房間是如此，人腦也是如此。如果中文房間這一整套系統在執行程式的時候，仍然不懂中文，人腦在執行相同程式的時候，為什麼就能懂中文呢？

你可能立刻想到，人之所以懂中文，跟人腦神經系統的運作是有密切關聯的，而這個關聯在中文房間故事裡根本無法呈現出來。瑟爾的故事忽略了這點。沒錯！這個故事並沒有考慮人腦神經狀態的運作和心智運作之



間的關聯。可是，算機功能論本來就不考慮大腦神經系統的運作，不是嗎！所以，不管圖林算機有多精緻的理論發展，對於研究人類心智本質這問題來說，算機功能論是必定失敗的。

瑟爾曾經將他的想法用推論的方式來表示：

心具有語意的性質。

圖林算機僅僅是語法的。

所以，圖林算機不可能用來解釋心的本質。

用我們之前的觀念來簡單摘述瑟爾的論點：圖林算機的計算單位雖然是表徵，但是對於表徵內容的考慮卻始終不曾出現在對於圖林算機的程式設計上。我們說過，圖林算機僅僅是一種操作符號的系統而已。所以僅僅以圖林算機來理解人的心智必定是不足的！

瑟爾的想法蠻有道理的。電腦能夠做出看似有思想的行為，但是僅從它的表現來看，並不足以表示電腦是有心



瑟爾 (John Searle) 英國牛津大學畢業，師從著名的哲學家奧斯丁。他在語言哲學有重大的貢獻。他認為對於語言進行哲學分析時，應該以「言說行動」（亦即人類說出句子的行為）作為分析的對象，而不是分析被說出的句子。這是有名的「言說行動論」。在 1980 年代他轉向「意向性」的研究，借用他自己的言說行動論來說明意向狀態。「中文房間論證」(Chinese Room Argument) 是他在心與認知哲學領域最知名的主張了。



智的。一臺電腦可以挑出你文章的英文錯字，但那只是因為它被內建了一套字典以及對照字母的程式，而不是它有思想、懂英文，不是它相信你拼錯字了，不是它想要幫你改正錯字。販賣機在接收到 10 元硬幣並給你一罐飲料的時候，它並不是想要跟你做生意。當它找給你 5 元硬幣的時候，它並不是真的知道你多給了 5 元，它並不是真的想要找你 5 元。販賣機只是在做設計者希望它做的事，它只是在執行程式，而不是像人一樣思考決策。這一切對於算機功能論來說，心智只等於那一套程式，那一套符號操作系統而已，至於符號所代表的意義，雖然通常寫程式的人心中有所考量，在執行程式的時候，並不考量符號的意義。尤其表徵的內容並不是內建的、並不是在設計圖林算機的時候，就可以「寫」進程式裡的。

或許你會提出這樣的疑問：既然算機操作的是符號或者表徵，符號是有意義或內容的，為什麼它的意義或內容沒有在執行程式中佔有地位呢？讓我們回顧販賣機的例子。

在前面販賣機的例子裡，那些硬幣就是表徵：有些圓形硬幣表徵了「5 元」，有些圓形硬幣則表徵了「10 元」。但是嚴格來說，自動販賣機並不「知道」你所投入的是 5 元還是 10 元（其實它也不知道它在進行買賣，而



且它也不知道它放出來的是飲料，而不是空盒子或者其他物品)。販賣機「知道」的其實是具備某種形狀、大小、重量、或者材質的物品。舉例來說，如果有一天政府將現在的 5 元硬幣改成不同的幣值，例如 5 角，但是不改變這種硬幣的大小、形狀、重量、材質等，這臺販賣機的程式仍然會照原來設計的方式來執行。只要投入那種硬幣兩個，你仍然會得到價值 10 元的飲料。那臺販賣機並不會辨別幣值的差異。如果政府改用不同形狀的硬幣，如正八邊形，來代表 5 元，那麼這臺販賣機的硬體部分必須作一番調整，但是它的圖林算機表仍然可以維持不變。瑟爾的批評就是指：雖然所使用的硬幣都有幣值，硬幣的幣值對於這臺販賣機的運作並沒有影響。

讓我再寫一個簡單的算機，並且故意用幾個奇怪的符號，來表達程式執行不受符號內容影響的意思：

	輸入: $\ominus$	輸入: $\oplus$
狀態一	(a) 1. 不做動作 2. 進入狀態二	(b) 1. 輸出: $\ominus$ 2. 停留在狀態一
狀態二	(c) 1. 輸出: $\ominus$ 2. 回到狀態一	(d) 1. 輸出: $\ominus$ 及 $\oplus$ 2. 回到狀態一

這個算機所處理的符號只有三種： $\ominus$ 、 $\oplus$  以及  $\ominus$ 。



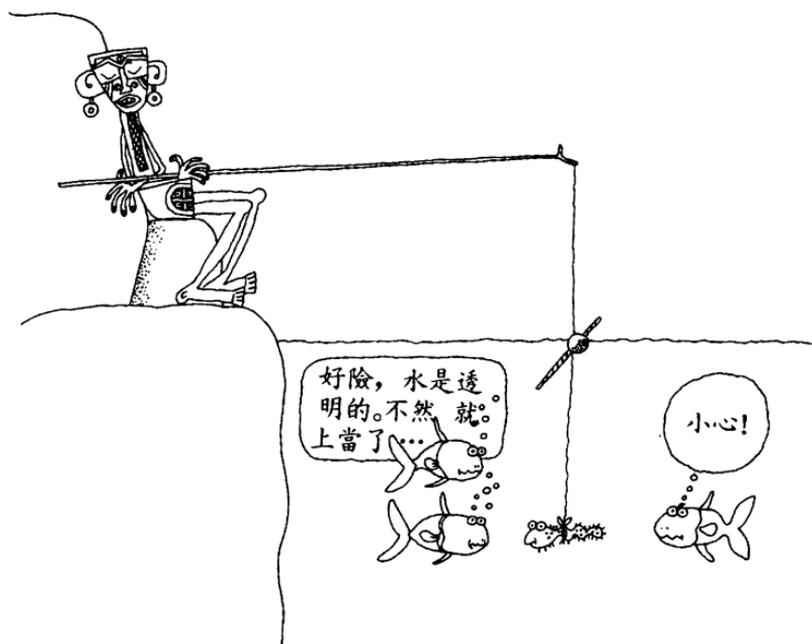
任何人即使不知道符號的意義，仍然能夠按照程式來執行指令。

你看得出來這臺電腦在執行什麼功能嗎？你大概很快就發現這臺電腦就是前面所介紹的販賣機。你的答案既對也錯。當你將「☉」詮釋為「5元」、「☽」詮釋為「10元」、「☼」詮釋為「一罐十元的飲料」時，這臺電腦就是前面介紹的販賣機。然而如果你將「☉」詮釋為「上網咖十小時」、「☽」詮釋為「寫作業一小時」、「☼」詮釋為「買手機一支」。這個電腦的意思是：如果你連續上網咖十小時，就讓你買一支手機；如果你上網咖十小時，再去寫一小時作業，就讓你再上網咖十小時，並且買手機一支；最後，如果你一開始就寫了一小時作業，就讓你買一支手機。

不管符號本身有沒有意義，電腦的運作是不管符號意義的。所以不懂中文的洋人是可以關在房間裡，讓那個系統讀完一本中文小說的。

那麼，電腦功能論有辦法解決中文房間論證所造成的困難嗎？如果對於表徵多一些瞭解，或許就能夠提出一套完整的關於「心智」是什麼的理論。當代哲學與認知科學大師佛德主張：在計算理論之外，再增補一套適當的表徵理論，就可以建立一套完整的理論來說明心智。讓我們再看下去！

# 6 人的思維





## 表徵、內容與意向性

我們曾經注意到心智的一個重要特徵是意向性；意向性表現在心理狀態的那種「指向」的特徵上。這個特徵在當代哲學界是用心理狀態的語意性質來加以說明的。「語意性質」指的是跟「意義」、「內容」有關的性質。例如我心中想著今天傍晚約朋友一起吃飯。我這個想法具有意向性，它指向「今天傍晚約朋友一起吃飯」這件事，這件事是我這想法的內容。當然我們已經知道，不是所有的心理狀態都有意向性這個特徵，「感質」這類心理狀態就沒有（痛、癢）。此外，心理狀態的意向性並不一定是指向外在世界的，也可以有心理狀態是指向內在心理世界的。例如我心中想著我非常喜歡布袋戲。我此時的想法指向的是我對於布袋戲的喜愛，「我對於布袋戲的喜愛」構成了我這想法的內容。

瑟爾在批評算機功能論時，指出心智是具有語意性質的，而算機的操作並不在乎這種性質，因為算機功能論僅僅著重在符號操作的層面，在程式的撰寫與執行中，符號的意義並沒有佔據重要的地位。瑟爾的說法等於是在批評：依照算機功能論製造出來的機器人，無法呈現意向性的特徵。既然算機操作的是表徵，如果對於表徵



多瞭解一些，應該能讓我們找出解決意向性問題的線索。

哲學家區分了表徵承載者以及表徵內容。表徵承載者跟它所表徵的內容並沒有必然的關聯。在販賣機的例子裡，那種形狀、大小、重量、以及那種材質的物體就是表徵承載者，它的幣值就是這種物體所表徵的內容。所以嚴格來說，「表徵」指的是某個物理的東西（表徵承載者）在某種運作方式之下，獲得了適當的內容，藉以代表某些事物或者某些事情。不同的物體可以用作相同的表徵，來代表相同的意義或內容。相同的物體也可以改變它所代表的意義或內容，而變成不同的表徵。

哲學家還區別了圖像式表徵以及語言式表徵。我們平常看到的照片、風景圖片、素描、簡圖、平面地圖、立體地圖、地球儀、模型……等，都屬於圖像式表徵。我們所使用的中文、英文等則是語言式表徵。語言式表徵不一定是用寫的，也可以是用說的。有些語言雖然沒有書寫的符號，但是有口語的表達，所以也是語言式表徵的一種。此外，人類刻意製造的語言也是語言式表徵，例如電腦語言、摩斯密碼等。童子軍或軍隊使用的旗語也算是一種語言式表徵；手語也是。還有，用數學符號表達的數學系統、用邏輯符號表達的邏輯系統等，都是語言式表徵。

圖像式表徵跟語言式表徵的差別在哪裡呢？基本上，



圖像式表徵跟被表徵的事物之間具有相似性。不過，這個說法還可以再細緻一點，圖像式表徵不一定是表徵一個個體，就像你的照片跟你自己一樣。圖像式表徵也可以表徵個體和個體之間的關係或者一群事物之間的結構。臺北市街道圖就是對於臺北市街道結構的圖像式表徵。其他像三維解析幾何也可以當作是各種物體空間分配方式的表徵，它用座標來代表各個物體之間的空間位置。這種表徵也是屬於圖像式的，因為座標反映了各個物體之間的空間結構。哲學家用數學裡的同構（亦即相同結構）概念來說明圖像式表徵跟被表徵事物之間的相似性。相對地，語言式表徵跟被表徵的事物之間並不相似。

細心的你應該發現有兩個關鍵問題遺漏了。第一個，剛剛提到了很多東西都是表徵，例如地圖和照片。難道這些東西都有意向性嗎？難道我們要承認這些東西都有心智嗎？第二個，當然也是最受重視的，關於心智的問題，亦即意向性的問題，如何跟表徵問題相連在一起？此外，心智所具有的表徵是圖像式的，還是語言式的？讓我先回答第一個問題，至於第二個問題留到下一節再說。





## 地圖與照片具有意向性嗎？

確實有很多東西都是表徵：照片、地圖、模型、語言……等都是。至於要不要承認它們具有意向性，並因此而具有心智，還有待斟酌。瑟爾指出人類具有原初意向性，也就是真正的意向性。但是地圖、照片和語言符號等，這些東西僅僅具有衍生意向性而已。他的意思是說，地圖和語言符號等之所以具有意向性，是由於我們這些具有真正意向性的人使用它們來傳達一些訊息或內容。我們是地圖和語言符號等表徵的製造人或者使用人。我們對於地圖和語言符號如何使用、要表達什麼訊息或內容有一些想法，我們建構了它們來傳達我們的想法。它們本身是沒有意向性、沒有內容可言的，它們的意向性和內容是來自於我們人類。

瑟爾其實還多說了一種，稱為疑似意向性。木偶是沒有生命的，當然也不會有意向性。但是我們可以假裝木偶是有意向性的，看木偶戲就好像看到了真正的人在思考、在說話、在做事一樣。皮影戲也是一個例子。這些事物本身不會有心智，不會展現意向性。另外更明顯的例子是我們將一些我們確定是沒有心智的物體當作是有意智的來看待。例如，有人心中也許想著：那塊石頭



一動也不動地待在樹下，是在履行它永不離開那棵樹的承諾。這個例子你大概會一笑置之，覺得無趣。不過，如果你仔細回想你所讀過的文學作品，有多少作品採用擬人法呢？如果你再仔細想想，我們的社會其實充滿了這種想法。石精樹靈不是只出現在電影裡的幻想而已。社會大眾在最瘋狂的時候，為了中彩券，幾乎是無所不拜的。這種瘋狂，從這裡的主題來看，已經是混淆了擬似意向性跟真實意向性。

現在我們可以來想想第二個問題了！

## 思維語言假設

### 心智的意向性是圖像式，還是語言式的？

我們大致上已經看到了表徵跟意向性的關係。在近代哲學史上，洛克、休姆等經驗論學派的哲學家認為，心理狀態跟外在世界的物體之間都具有表徵的關係，而且這種表徵是圖像式的，例如我的「樹」的概念是我心中的一個樹的圖形。這個想法在當代已經失去了吸引力。在當代主要是針對具有意向性的那些心理狀態，經由表徵理論來瞭解它們所具有的內容。這一點跟算機功能論



的發展也有密不可分的關聯，因為算機功能論就是主張心智是一套符號操作系統，所以表徵理論有補足算機功能論的價值。這也是羅格斯大學哲學系教授佛德的想法。佛德的想法在哲學界和認知科學界通稱為心理表徵論，不過他的理論有另外一個更為人知的名稱：思維語言假設。

思維語言假設，顧名思義，主張人的思維是語言式的，也就是以語言式的表徵作為操作的對象，並以語言式的、邏輯的方式來進行操作。不過，佛德的理論只侷限於命題態度這類心理狀態，對於感質之類心理狀態並不適用。這是由於佛德一心希望對於人的各種思考、判斷、信念、推想等認知層面的心理結構（簡稱認知構造）提出理論的說明。佛德並沒有著力在心智的非認知層面，例如喜怒哀樂等的研究。這部分的研究是由其他哲學家開發的。

什麼是語言式表徵？我們在上一節看到圖像式表徵跟被表徵事物之間具有相似性（或者同構性），語言式表徵並沒有這個特徵。語言式表徵最重要的特徵有兩個：它是「原子式」的，以及它具有真假值，也就是它或者為真，或者為假。圖像式表徵並沒有這兩個特徵。讓我用中文作例子來解說。

在中文裡，「有些玫瑰是紅色的」是一個真的句子。



我們通常說「玫瑰」和「紅色的」是原子詞，並造出「玫瑰是紅色的」這個句子；「玫瑰」和「紅色的」這兩個語詞本來就有自己的意義，並進而構造出這個句子的意義。此外，「有些玫瑰是紅色的」還可以用一些連接詞跟其他句子合起來，以造出複雜的句子。例如「有些玫瑰是紅色的，但是玉蘭花都是白色的」使用了「但是」這個連接詞來連接「有些玫瑰是紅色的」以及「玉蘭花都是白色的」這兩個句子。

哲學家說，語言式的表徵具有這種稱為語意組構性的特徵：複雜句的意義是由簡單句的意義組合而構造出來的。人類的認知或思維也具有這種特徵，複雜的思維是由簡單思維所組合出來的。佛德之所以提出思維語言假設，就是為了要解釋中文、英文或者其他語言所具有的這種語意組構的特徵。他的思考是這樣的：語言（中

佛德 (Jerry A. Fodor) 20 世紀後半期深具影響力的哲學家之一。他在心與認知哲學、心理學哲學、以及認知科學的發展上所做的貢獻無遠弗屆，他的「思維語言假設」更是這個領域必定要知道的哲學學說。他一手建立古典學派，他關於「模組」概念的發揮影響了認知科學以及演化心理學對於認知構造的瞭解，他的先天論學說承襲當代語言學大師瓊姆斯基的「普遍文法」理論，對於「概念」的理解有重大影響。佛德著作繁多，影響重大者包括《思維語言假設》、《心理表徵：論文選》、《模組性》、《心理語意學》等。





文、英文等) 具有語意組構性。語言表達思想，所以思想最好是具有語意組構性，否則難以瞭解為什麼語言具有語意組構性。思想的語意組構性從哪裡來呢？表徵理論要能充分解釋思想的語意組構性，莫過於承認心理表徵也是語言式的，而不是圖像式的。佛德之提出思維語言假設就是來說明這點的。

佛德認為思維是一套由心理表徵構成的系統，而且這些心理表徵都是語言式的。他主張在這套系統裡有概念表徵（相當於中文「有些玫瑰」、「紅色」等語詞），以及由概念表徵組合出來的命題表徵（相當於中文「有些玫瑰是紅色的」之類的句子）。命題表徵都有真假值，而且有些命題表徵是原子的，其他的命題表徵則是由原子命題表徵組合出來的複雜表徵。除此之外，沒有任何一個表徵能夠完整表徵出被表徵物的所有面向。「有些玫瑰是紅色的」這個句子只表徵了玫瑰的顏色，並沒有表徵到玫瑰的其他性質。最後，佛德還主張所有原子式的概念表徵都是天生的、與生俱有的。（不過他不曾明確列出有哪些原子式的概念表徵。）佛德的思維語言假設加上算機功能論，這整套學說影響了 20 世紀認知科學的發展，被稱為古典論或古典學派。

我在前面曾經提到哲學家對於「表徵承載者」以及「表徵內容」的區別，而且相同的表徵內容可以由不同



的事物來承載。在佛德的學說裡，人的大腦狀態就是心理表徵的承載者。因此，以不同於人腦的材質做出來的機器人也能擁有表徵。不管是外星人還是機器人，即使大腦材質不同、腦結構不同，仍然能夠具有表徵內容，因此佛德的學說可以說明外星人和機器人的認知構造。

佛德的思維語言假設跟學術界任何學說一樣，都會面臨挑戰，也都會有競爭的學說出現。不過，不論思維語言假設是不是被接受，以表徵理論來瞭解心理世界受到當代絕大多數哲學家的支持；在當代認知科學界，「表徵」跟「計算」一樣，構成了它的核心概念。用心理表徵理論來說明思維的內容，並藉以說明思維的意向性，確實是在心智的研究上跨越了一大步。畢竟，在之前我們就注意到心智的重要特徵之一是它的意向性。如果我們瞭解意向性，我們對於心智就多了一層認識。

## 孔明與諸葛亮

正如你預期的，事情沒有我們想像中的簡單。不論你是不是接受佛德的思維語言假設，甚至不論你是不是接受心理表徵理論，對於思維內容的研究仍然面臨一些重大的關鍵難題。在這一節和下一節我要特別提出兩個關鍵難題。



我們首先遭遇的難題跟指涉隱蔽性有關。什麼意思呢？假設張三對於「孔明」這個名字很熟，但不知道為什麼他竟然完全沒有聽過「諸葛亮」這個名字。他聽過很多關於孔明的事蹟，不論這些事蹟是真是假。孔明設下那有名的空城計，騙倒了司馬懿；孔明是劉備最重要的軍師；孔明娶了一個有才德但相貌很醜的老婆；孔明發明了木馬流牛；孔明草船借箭……等。根據這個假設，如果你問張三：「諸葛亮有沒有結過婚？」由於張三根本不曉得諸葛亮是哪號人物，他沒有辦法回答你的問題，儘管張三相信孔明有娶老婆。這表示在張三心中，「孔明有娶老婆」跟「諸葛亮有娶老婆」是兩回事。

再舉一個例子。根據剛剛的假設，我們可以明確地說：「張三相信孔明設下空城計」。然而你有多大信心敢說：「張三相信諸葛亮設下空城計」？如果你去問張三：「你相信諸葛亮設下空城計嗎」？張三是會否認的。這表示對張三來說，「相信孔明設下空城計」跟「相信諸葛亮設下空城計」是兩件不同的信念。可是，諸葛亮明明就是孔明，張三相信的是同一個人所做的同一件事（設下空城計）。為什麼張三會產生「那是兩件不同想法」的情形？

也許你還不太服氣，看不出這裡究竟有什麼哲學問題。讓我們換一個例子。在佛洛伊德精神分析學裡有一



佛洛伊德 (Sigmund Freud, 1856-1939) 奧地利內科醫師、神經學者及精神分析學派的創始人。以自由聯想的方式推展精神分析法，在神經學、心理分析研究各方面，都有創新的理論。著有《夢的分析》、《圖騰與禁忌》、《自我的分析》等書。

個很有名的伊底帕斯情結，也就是「戀母情結」。這是來自於古希臘的一個故事。伊底帕斯在意外的情形下愛上了一位美女雅克斯姐，並準備跟她舉行婚禮。他相信他會娶雅克斯姐為妻。但誰也沒料想到雅克斯姐竟然是伊底帕斯的親生母親。在伊底帕斯相信他會娶雅克斯姐為妻時，他並不相信他要娶他自己的親生母親，否則他也不會大費周章，安排盛大的婚禮，也不會在他發現他想娶的人竟然是自己的親生母親後，發生了悲劇。由於人的想法影響人的行為，如果「伊底帕斯會娶雅克斯姐為妻」這個想法跟「伊底帕斯會娶他自己的親生母親為妻」這個想法是一樣的，就不應該會引起伊底帕斯不同的行為、不同的判斷、不同的心情。相同的想法在相同條件下應該引起相同的行為；反過來說，在相同條件下，如果行為是不同的，想必是由於導致行為的想法不同的緣故。

然而伊底帕斯難道不是對同一個人想同一件事嗎？難道張三不是針對同一個人想同一件事嗎？一方面因為



引起不同的行為、不同的思考和判斷，我們承認兩者應該是不同的想法；另一方面又不得不承認兩者是對於完全相同的事情的想法，所以兩者應該是相同的。我們面臨一種「掙扎」，不知道究竟應該斷定張三擁有的是兩個不同的想法，還是仍然只是一個想法而已；不知道究竟應該主張伊底帕斯的兩個想法是相同的，還是不同的。哲學家說，這種掙扎是由「指涉隱蔽」的現象引起的，所以這個掙扎被稱為指涉隱蔽難題。

什麼是指涉隱蔽？「孔明」這個名字的指涉就是孔明這個人，和「諸葛亮」這個名字的指涉都是同樣的那個人。指涉隱蔽的意思是說，在某些情形下，名字的指涉會被遮住，讓人掌握不到。人的思維就有這種情形。如同張三的故事一樣。雖然孔明和諸葛亮兩個名字有相同的指涉（同樣那個人），但是在張三的思維中，兩個名字的指涉被遮住了，張三雖然相信孔明發明了木馬流牛，他並不相信諸葛亮發明了木馬流牛。這種指涉隱蔽的現象在伊底帕斯的故事也是一樣的。

照這樣來看，儘管張三的兩個想法（亦即「孔明有老婆」以及「諸葛亮有老婆」）都是針對同一個人的婚姻狀況的想法，由於人的思想有指涉隱蔽的現象，張三第一個想法所想的事情，就不應該跟他第二個想法所想的事情一樣。哲學家提出的一個解決之道是：這兩個想法



具備不同的內容。在「孔明與諸葛亮」的故事裡，雖然張三的兩個想法是針對同一個人，這個人卻是以不同的方式呈現在張三的心中。換成表徵理論的話來說，張三對於同一個人作了不同的表徵（別忘了，沒有任何表徵能表徵被表徵物的所有面向，每個表徵都是表徵到被表徵物的某個面向而已）。關於孔明的想法跟關於諸葛亮的想法，心理表徵理論以「兩者具有不同的內容」來回應指涉隱蔽難題。

## 雙生地球

在對於人類思想提出理論時，「孔明與諸葛亮」的故事所造成的是指涉隱蔽難題。雙生地球的故事卻是從反方向來製造理論困難。這個故事是哲學家帕南提出來的。他要我們作一些想像（我作了一些修補）：

假想一個新的宇宙。在這新的宇宙裡所有一切跟真實的宇宙非常相似，它有銀河，銀河中有太陽系，太陽系中也有地球。尤其這個地球跟真實的地球幾乎一模一樣。它的自然環境跟真實地球是一樣的；它的各個人種、各個國家的地理分佈都是一樣的；它的人文歷史的發展是一樣的；



它的所有語言，不論是中文、英文還是希臘文，也都一樣。

除了一點：新地球的太平洋裡，水的化學結構跟真實地球的水的化學結構不同，其實它的所有海洋、湖泊、河流的水的化學結構跟真實地球的水的化學結構不同。真實地球上的水化學結構是  $H_2O$ ，這個新地球的水化學結構則很複雜，為了方便，姑且稱為 XYZ。儘管如此，所有真實地球上水的地方，新地球上也都一模一樣，包括天上下的雨、人體內的水的比例、仙人掌所貯存的水……等，通通都一樣。而且新地球的水跟真實地球的水一樣，都是無色無臭無味的流動液體，都是人以及所有動植物賴以為生的必要因素。

讓我們將這個假想的新地球稱為「雙生地球」。也許你覺得這個故事不太對，怎麼可能兩種化學結構不同的物質卻有相同的表面特徵，例如無色無臭無味？又怎麼可能會跟其他任何物質都出現相同的反應？人體血液內有  $H_2O$ ，雙生地球人體血液內有 XYZ，兩者怎麼會有相同的生理功能？不過我們可以修改故事，以避免你這種質疑，而仍然維持故事的用意。例如故事可以進一步這樣假想：XYZ 這種物質具備所有  $H_2O$  的特徵以及化學性質，但是 XYZ



還有一些特徵和化學性質是  $H_2O$  沒有的，只是這些特徵在地球和雙生地球都呈現不出來。

藉著雙生地球的故事，帕南想問的問題是：在雙生地球上的人使用語言的時候，語言的意義有沒有發生變化？既然在我們的故事中說真實地球跟雙生地球之間只有水的化學成分不同，就用水作例子，來繼續這個故事：

地球上有一位王五，沒學過化學，不知道水的化學成分（我們甚至可以更極端地假想沒有任何人知道水的化學成分）。他在下了課之後覺得口渴，就喝了杯水。他心中有很多想法，其中包括：「水是透明的。」

在雙生地球也有一個跟真實地球的王五一模一樣的人，為了方便起見，讓我們叫他「雙生王五」。雙生王五沒學過化學，不知道水的化學成分。雙生王五在雙生地球下了課之後也覺得口渴，也喝了杯水。他心中也有很多想法，其中一個是：「水是透明的。」

帕南質問：他們兩個在想相同的事情嗎？究竟地球王五跟雙生王五「水是透明的」這兩個想法是一樣的，還是不一樣的？它們有沒有相同的內容？

地球上的王五知道水是無色無臭無味的，他知道水



是液態的，他知道油會浮在水上，他口渴時會喝水，他澆花時會用水，雨不大時他喜歡淋一身有點濕卻不太濕的感覺……。地球上的王五用中文「水」這個字來說他口渴時喝的東西、澆花時用的東西……等。雙生地球上的雙生王五也是一樣，他也對於雙生地球上的水有相同的知識，尤其他也是用中文「水」這個字來說他口渴時喝的東西、澆花時用的東西……等。「水」這個字的用法對他們沒有差別，他們的行為反應也一樣沒有差別。





照這樣來看，地球王五跟雙生王五關於水的想法應該是一樣的，兩者應該有相同的內容，因為他們兩人處於相同的情境，接受相同的環境刺激，也做出相同的行為反應。如果那兩個關於水的想法是不一樣的，如果那兩個想法的內容是不同的，那麼那兩個關於水的想法不應該對兩人都是引起相同的行為。在「孔明與諸葛亮」的故事裡，我們提到：相同的想法在相同條件下應該引起相同的行為；反過來說，在相同條件下，如果行為是不同的，想必是由於導致行為的想法不同的緣故。所以王五「水是透明的」的想法跟雙生王五「水是透明的」的想法應該是一樣的，應該有相同的內容。

然而事情沒那麼簡單。帕南自己認為兩者是不一樣的想法，兩者的內容不同，因為思想的內容是由思想所及的外在對象來決定的。如果你心中想著：「孔明是三國時代的人」，我心中想著：「劉備是三國時代的人。」我們兩人的想法當然是不一樣的，因為我們兩人所想的對象不同，一個是孔明，一個是劉備。同樣地，根據故事，地球上的王五所使用的「水」這個中文字指的是  $H_2O$  這種物質；雙生地球上的雙生王五所使用的「水」這個中文字指的是 XYZ 這種物質。既然如此，地球王五「水是透明的」的想法跟雙生王五「水是透明的」的想法應該也是不一樣的，它們的內容應該是不同的，因為他們所



想的對象不同，一個是具備  $H_2O$  這種化學結構的物質，一個是具備 XYZ 這種化學結構的物質。所以，即使  $H_2O$  呈現給王五的以及 XYZ 呈現給雙生王五的是一模一樣的，由於兩者是不同的物質，王五和雙生王五的想法是不同的。

我們可以將雙生地球的故事說得更複雜。設想第二個新宇宙，在那裡也有地球跟雙生地球。地球王五在地球上對著那杯水想著：「水是透明的。」接著，地球王五坐太空船抵達了雙生地球，在那裡他對著那杯水又想著：「水是透明的」（王五怎麼老是想著水是透明的這件事？）這個情形是同一個人的兩次想法，用中文表達出來是完全一樣的。那麼，這兩次想法是一樣的嗎？它們具有相同的內容嗎？帕南主張這時候地球王五在地球的想法跟他在雙生地球的想法是一樣的，它們具有相同的內容。這是因為地球王五「水是透明的」的想法其內容在地球上就已經被決定了，是關於  $H_2O$  這種液體所具有的透明性。既然這個思想的內容已經如此決定了，不論地球王五到哪裡，他「水是透明的」的想法其內容並沒有改變，即使他到了一個事實上不是  $H_2O$  而是 XYZ 的地方。同樣地，雙生王五在雙生地球「水是透明的」的想法，其內容是關於 XYZ 的。如果他也坐太空船到地球旅行，他在地球上「水是透明的」的想法，其內容仍然是關於 XYZ



的。

對於雙生地球故事討論之熱烈，恐怕在當代哲學界是很少見的。雖然還有一些相當重要的爭議，帕南的主張得到很大的支持。

如果你重新想想「孔明與諸葛亮」的故事以及雙生地球的故事，你會發現我們可能已經身陷困境。我們對於這兩個故事所提的主張顯然是衝突的。

「孔明與諸葛亮」的故事指出思想內容有指涉隱蔽的現象。在那故事中，我們認為關於孔明的想法跟關於諸葛亮的想法是不同的，儘管這兩個想法是針對同一個對象。我們是以呈現給認知主體的方式來決定思想內容的同異。然而這個主張在雙生地球的故事裡遇到衝突。 $H_2O$  的呈現方式跟 XYZ 的呈現方式是一樣的，所以我們本來應該主張那兩個「水是透明的」的想法是相同的。但是我們發現剛好相反。那兩個「水是透明的」的想法是不同的，因為那兩個想法是針對不同的對象。

另一方面，假設我們從思想所及的對象來決定思想的內容。難道我們要主張在「孔明與諸葛亮」的故事中，「孔明有老婆」的想法跟「諸葛亮有老婆」的想法是一樣的嗎？難道我們要承認在伊底帕斯的故事中，「伊底帕斯會娶雅克斯妲為妻」的想法跟「伊底帕斯會娶他自己的親生母親為妻」的想法是一樣的嗎？這跟我們之前的



回應又互相衝突。我們如何回應我們之前主張兩者是不一樣的理由？

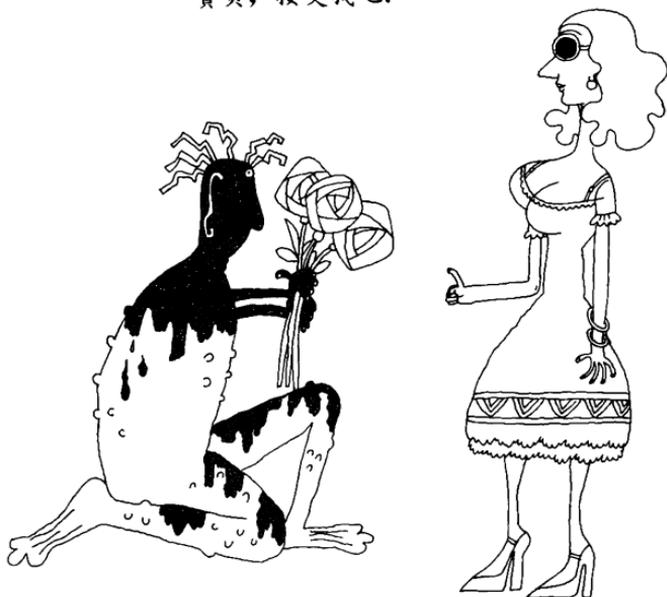
我們所面臨的困境是：我們無法決定要以「對象呈現給主體的方式」來決定兩個想法是相同的，還是不同的；還是要以「思想所及的對象」來決定兩個想法是相同的，還是不同的。如果採用第一種方式來作決定，則對於「雙生地球」故事裡的兩個思想會作出不當的判斷。如果採用第二種方式來作決定，則會對「孔明與諸葛亮」故事中的兩個思想作出不當判斷。

意向性的問題跟思想內容的問題是一體兩面的，如果能對於意向性或者思想內容有清楚的瞭解，對於我們心與認知的研究無疑跨出了一大步。但是我們還面臨許多難題需要克服，這工作就留給有心人吧！



# 7 感 質

寶貝，接受我吧！





我們已經看過物理論的兩個重要主張：邏輯行為主義以及心腦同一論；我們也看到因果角色功能論以及算機功能論（加上表徵理論）對於心（心智）的說法。雖然嚴格來說，算機功能論不是物理論，但是跟物理論是相容的，而且當代哲學家和認知科學家也都在接受功能論的同時，走向物理論的立場。物理論確實在當代佔有絕對多數的優勢。

不過，這不表示物理論的立場沒有受到挑戰。當代一些哲學家分別提出了幾個問題來質疑甚至反駁功能論和物理論。在這一章我會將這些質疑和反駁作一些解說，主要包括感質對調的問題、感質缺如的問題、無識人（沼澤人）的問題、以及黑白瑪麗的故事。前面兩個問題是針對功能論提出來的，後面兩個問題是針對物理論提出來的。最後，物理論最大的困難，也是研究心與認知的哲學家 and 科學家一直希望突破的，恐怕是要如何理解意識的主體性（我會在這一章最後一節作一些解說）。



## 我看到的是綠色的玫瑰

### 反對功能論的理由之一

功能論遇到的第一個問題稱為「感質對調」。感質指的就是我們的喜怒哀樂七情六慾等，那些不具有意向性的心理狀態。在 17 世紀英國哲學家洛克提出了感質對調的可能性，來談論色彩知覺的光譜。我用故事將這個可能性作一番說明：

張三是個跟我們沒有什麼兩樣的正常人。他看到紅色玫瑰的時候，會產生「紅色」的視覺經驗；當他看到玫瑰花的綠色葉子時，他會產生「綠色」的視覺經驗。此外，張三當然懂中文。他會用「紅色」這個中文字來描述他所看見的玫瑰顏色，以及其他東西的相似顏色，例如血液、朱槿等；他會用「綠色」這個中文字來描述他所看見的玫瑰葉子的顏色，以及其他東西的相似顏色，例如夏天地上的草皮。他當然會有「玫瑰花是紅色的」、「樹葉是綠色的」之類的想法。



有位王五，跟張三一樣，他也會用「紅色」這個中文字來描述他所看見的玫瑰顏色，以及其他東西的相似顏色，例如血液、朱槿等；他也會用「綠色」這個中文字來描述他所看見的玫瑰葉子的顏色，以及其他東西的相似顏色，例如夏天地上的草皮。他當然會有「玫瑰花是紅色的」、「樹葉是綠色的」之類的想法。

不過，王五跟張三有一點很大的不同：王五看到樹葉時所產生的色彩經驗居然跟張三看到玫瑰時所產生的色彩經驗是一樣的；而王五看到玫瑰時所產生的色彩經驗居然跟張三看到樹葉時所產生的色彩經驗是一樣的。也就是說，他們兩人的「紅」、「綠」色彩知覺的經驗是對調的。

洛克認為這個故事是有可能的，當代許多哲學家也同意他。但是這樣一來，問題就出現了。別忘了功能論的主張是說：任何心理狀態都等於某種功能狀態；功能狀態或者等同於計算狀態，或者是從因果角色來定義。因此，張三看到玫瑰所產生的色彩知覺經驗，應該跟王五看到玫瑰所產生的色彩知覺經驗是一樣的；他看到樹葉所產生的色彩知覺經驗，應該也跟王五看到樹葉而產生的色彩知覺經驗是一樣的。

但是這個故事卻告訴我們，即使是相同的環境刺激，



仍然能造成不同的色彩知覺經驗。張三和王五的色彩知覺經驗是對調的。而且別忘了，他們兩人對於相同視覺刺激所做的行為反應並沒有不同，他們所用的語言也沒有差別。張三用「紅色」這個中文字來描述他所看到的玫瑰顏色，他用「綠色」這個中文字來描述他所看到的樹葉顏色。在這方面，王五的語言表現跟張三是一模一樣的。如果你接受這個感質對調的故事，那麼你也就是認為功能論是錯誤的。

你大概很想要抱怨了。講一個感質對調的故事，純粹憑著故事的可能性就來提出反駁，這種作法太過容易了。按照這種作法，對於任何理論只要編一個奇怪的故事就可以反駁了嗎？這抱怨很有道理，也的確觸及哲學思考的方法問題。並不是所有編出來的故事，僅僅憑著一個可能性，就可以成為一個主張或者反駁一個理論。

哲學界將這類所謂的故事稱為思考實驗。在 20 世紀的英美哲學界，思考實驗是很常見、很重要的一種哲學思考方式。我們其實已經看過很多思考實驗了：「孔明與諸葛亮」、「雙生地球」、「中文房間」、「完美偽裝者」、以及「超級斯巴達人」。這些故事都是哲學家的思考實驗。每個思考實驗都是哲學家用來表達某個哲學觀點的一種方式，或者是用來反對某個哲學觀點的一種方式。因此，一個思考實驗的好壞，涉及到這個故事是不是表達了說



故事的哲學家想要表達的觀點，或者是不是表達了他反對某個觀點的用意。

以目前這個感質對調的故事來說，確實憑著故事的可能性就來提出反駁，是不夠的。還好，這個故事針對的是功能論，而不是心腦同一論。假設這個故事針對的是心腦同一論。當王五看到玫瑰時，他看到玫瑰而產生的色彩知覺經驗，跟張三那個色彩知覺經驗，是屬於同一類型的心理狀態。依據心腦同一論，張三看到玫瑰所產生的那類型的色彩知覺經驗，會等同於某種類型的腦神經狀態。由於王五跟張三同樣都是人類，所以他看到



玫瑰而產生的那類型的色彩知覺經驗跟張三所處的腦神經狀態是同一類型的。既是如此，心腦同一論不一定要接受感質對調的可能性。所以，僅僅憑著這個故事本身未必能反駁心腦同一論。

不過功能論就很難說了。功能論是用功能角色來定義心理狀態的。張三看到玫瑰而產生的色彩知覺經驗，以及王五看到玫瑰而產生的色彩知覺經驗，即使兩者不同，在功能論裡，兩者仍然可以扮演相同的功能角色。如此一來，功能論無法排除感質對調的情形。然而一旦無法排除感質對調的可能性，我們就沒有辦法用功能角色來定義心理狀態了。張三和王五在相同環境下，



接受相同的刺激，產生相同的行為，就連語言表現也無法分辨。但是他們卻擁有不同的色彩知覺經驗。感質對調的可能性因此威脅了功能論。

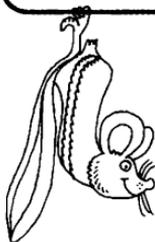
本書包括本章在內，都有很多故事（思考實驗）。這些故事的用意都在於凸顯一些我們平常願意接受的想法，或者試圖指出某個理論應該涵蓋一些哲學觀點，但卻沒有涵蓋到。這是為什麼思考實驗在哲學研究方法上是很重要的。我們也將從這個角度來理解接下來的其他故事。

你你你……  
每次都這樣！  
難道你不知道  
我最討厭  
黑色的玫瑰？

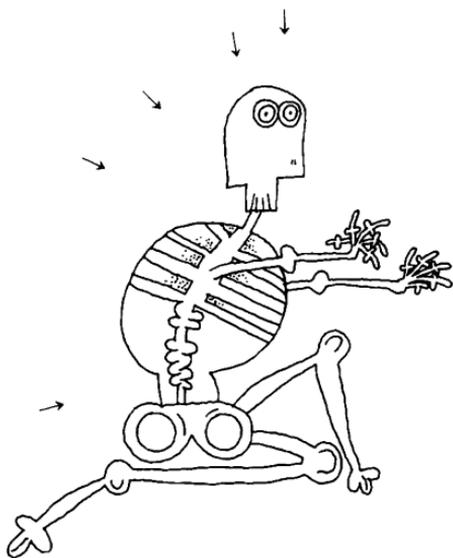




洛克 (John Locke, 1632-1704) 牛津大學畢業，是近代哲學史上經驗論的第一代表人物，與笛卡爾的理性論相抗衡。經驗論主張所有的知識都來自於感官知覺的運作，對於當代哲學的發展仍然影響重大，例如當代的邏輯實證論就是吸收了經驗論的哲學，來探討科學知識的。洛克在知識論和政治哲學都有極大的貢獻。他的哲學著作有《人類悟性論》、《政府二論》等。



黑色的花？  
可是我只看到白色的骷髏地！





## 中國頭

### 反對功能論的理由之二

「中國頭」是個很奇怪的故事。紐約大學哲學系教授布拉克在 1980 年提出了這個故事來指出功能論沒有辦法說明人的感質。布拉克的想法是這樣的：既然按照功能論的說法，每一個心理狀態都是由它的功能角色（因果角色或者計算角色）來定義的，既然心智就是這樣一種功能系統或計算系統，那麼任何滿足這種功能系統或計算系統描述的物體應該是有心智的，應該能思考、能喜怒哀樂。但是他認為這裡存有可待質疑的空間。他要我們想像這樣的一個情景：

假設有一天我們寫出了「痛」的程式（或者找到了「痛」的功能角色）。

由於中國領土龐大，有十幾億人口，我們就來想像，在中國領土上每一個中國人都發給一臺收發機。感覺上，在中國領土上的這一群中國人，整體來看，好像構成了一個頭，所以我們用「中



國頭」來稱呼這個龐大的「東西」。同時，我們使用人造衛星來控制資訊的收發。或許要一千多人的收發機收到某些資訊才能代表程式裡的一個符號，或許要五千多人的收發機收到某些資訊，才能表示程式裡的另一個符號。由於中國人口眾多，我們應該不必太擔心人會不會不夠用。現在，我們既然有這一套關於「痛」的程式，我們就可以使用十幾億配給了收發機的中國人來執行程式。理論上，當「痛」的程式執行完畢時，這個「中國頭」就會出現痛覺。中國頭會覺得痛！

同樣的故事也可以拿來談中國頭的喜怒哀樂等感覺，乃至於中國頭的色彩知覺經驗、中國頭的嗅覺經驗……等。

你同意嗎？你會覺得這個「中國頭」真的有痛覺、真的有喜怒哀樂嗎？既然按照功能論的說法，「痛」的因果角色是可以定義出來的，「痛」的程式是可以寫出來的，如果這個故事一切按照功能論的說法，讓中國頭扮演「痛」的因果角色或者執行「痛」的程式，我們必須承認中國頭是會覺得痛的。但是，十幾億人口在那麼廣大的一片土地上，使用收發機作一些收發資訊的動作，就使得這整體會有痛覺，這太不可思議了。一般人很難相信這樣的「東西」會有痛覺。布拉克認為「中國頭」是



不會有痛覺的，更不會有任何其它的感質。他將這個故事稱為感質缺如的問題。僅僅執行程式，僅僅考量因果角色或計算角色，不足以說明人的感質，人的喜怒哀樂七情六慾。

你可能想到中國頭太龐大了，人數太多了，這套「痛」的程式要花很久的時間才能跑完，所以我們才會覺得它沒有痛覺。但這純粹是技術問題，只要將來技術獲得改善，就可以縮減執行程式的時間。更何況我們是從理論的角度來思考功能論的問題，所以技術問題是不相干的。其實布拉克也知道有人會提出這種疑慮，所以他另外想了一個類似中國頭的例子，但是這次他將所有的東西都縮小。這次他假設我們發明了很小的收發機，小到肉眼看不到。我們將十幾億個這種細微收發機放在一個跟人的頭大小相近的容器內（別忘了，我們現在已經有奈米技術）。然後我們來執行這套「痛」的程式。布拉克問：這個容器會覺得痛嗎？這個容器會有喜怒哀樂嗎？他還是認為不會。所以他主張依據功能論造出來的東西缺乏感質，功能論無法說明人的感質，人的喜怒哀樂七情六慾。

我們要注意一點：根據布拉克「感質缺如」的說法，功能論不是不完整的理論，而是錯誤的理論。這是因為功能論號稱是說明所有心理狀態的理論，任何心理狀態都是某種功能狀態（由因果角色或計算角色定義的狀



態)。但是如果中國頭的例子是有道理的，感質這類心理狀態不會等於任何功能狀態。所以，功能論是錯誤的。

或許我們有一個比較保守的作法。我們保留功能論對於思想、判斷、信念等意向狀態的說明，至於感質之類的心理狀態則必須另外發展新的理論。畢竟，感質之類的心理狀態既然跟思想、判斷等心理狀態有那麼大的差異，對於這兩大類心理狀態分別提出理論說明，似乎也不是不可行的。若是如此，「感質缺如」的問題至多指出功能論是不完整的理論，而我們剩下的問題就是另外發展一個新理論來說明感質是什麼，而且我們還要提出一套能整合這兩套理論的學說。這個關於感質的新理論會是什麼呢？這個整合的說法會是什麼呢？到目前為止，並沒有任何哲學家 and 認知科學家對於這個想法作過嘗試。也許你可以試試看！



布拉克 (Ned Block) 紐約大學哲學系教授，當代心與認知哲學大師之一。對於功能論提出許多深入的批評，引起不少的討論與迴響。在心理語意學方面提倡所謂的「概念角色功能論」，與佛德的「語意原子論」相抗衡。



## 沼澤人

### 反對物理論的理由之一



現在我們要思考的故事相當簡短，但卻是非常地荒誕不經。（你大概以為哲學家都是寫科幻小說出身的!?) 這個故事最早是由哲學家德威笙提出來的，後來這類故事陸陸續續用不同方式呈現出來。這個故事大致上是這樣的：

假想在某個人煙稀少的沼澤地，有一天風雨交加、雷霆交作。不知什麼緣故，在一陣閃電雷劈之後，從沼澤裡冒出了一個人。這個人當然不是妖怪，而是貨真價實的人。讓人訝異的是，這個沼澤人竟然跟你認識的張三在生理、物理方面是絕對完全相同的。

哲學家質問的是：這個沼澤人有心智嗎？有意識嗎？如果你覺得沼澤人的故事太科幻了，實在很難思考其中的問題，我們可以換一個聽起來比較像科學的故事。複製羊不僅是可能而已，它已經是個存在的事實。複製人在



目前科學技術來看，應該是指日可待的。當然，目前的技術複製出羊或者人的過程還很冗長、複雜，也不足以凸顯哲學家所思考的要點。哲學家所假想的複製人是跟某個有意識的真人（例如李四）在微觀物理的層面來看，是完全相同的。所以讓我們進一步假想，有一天科技已經發達到能夠製造出在每一個物理層面上都跟李四一模一樣的複製李四。只不過真的李四是有感質、有意識的，複製李四則是沒有感質、沒有意識的。哲學家要問的是：我們要接受「複製李四」是沒有感質的假設嗎？我們在理論上可以接受這種複製李四嗎？我們在理論上能夠接受沼澤人嗎？我們在理論上能夠接受一個沒有意識，但生理和物理方面跟有意識的真人一模一樣的個體嗎？哲學家將這種人稱為無識人，所以簡單來說：我們願意承認無識人嗎？

對功能論來說，如果兩個個體在物理層面上完全相同，他們在功能論所描述的層面上也會完全相同。既然這個沼澤人在生理和物理層面上跟真的李四一模一樣，所以只要功能論的學說能夠說明真李四的感質或意識，功能論就會承認沼澤人也是有感質、有意識的。當然，前面中國頭的故事對於這點已經提出了質疑。

沼澤人的故事對物理理論來說，情形有點複雜。這個故事引起了關於物理理論的重要爭議。如果接受沼澤人的



假設，就表示感質或意識不會等同於任何大腦狀態。因為真李四跟沼澤人在生理和物理上是一模一樣的，卻比沼澤人多了感質或意識，所以李四的感質或意識不會等同於任何大腦狀態。不但如此，如果我們接受沼澤人的假設，我們就不能主張大腦的運作一定會產生感質或意識，因為沼澤人的大腦跟真李四是一樣的，但真李四有感質或意識，沼澤人卻沒有。

你可能很快想到，心腦同一論這派物理論不是已經被功能論取代了嗎？我們又何必再考慮物理論是不是正確的問題？有一種物理論跟心腦同一論採取不一樣的立場。這派物理論稱為非化約物理論。雖然這個理論不主張心腦之間的等同關係，但仍強調心智和大腦之間有某種緊密的關聯。非化約物理論說，兩個個體只要處於相同的物理狀態（如腦神經狀態），則他們必定處於相同的心理狀態（用他們的術語來講，心智是隨附在腦之上的）。使用「隨附」這個概念，物理論就可以否認無識人了。因為依據假設，無識人的所有物理狀態跟李四是完全一樣的，因此，依據「隨附」概念，無識人必定跟李四處於完全相同的心理狀態。但是依據假設，無識人沒有感質，所以我們推論出一個矛盾。

總結來說，只要你接受無識人的假設，不管哪一種物理論，包括心腦同一論和非化約物理論，都是錯誤的。



那麼，我們要接受沼澤人的假設嗎？我們已經遇到過類似的哲學思考方法的問題，僅僅提出一個可能性是不足以反駁一個理論的，必須這個可能性本身含有雙方原本就同意接受的要點，以至於該理論跟這可能性無法相容；或者這個可能性能夠指出被反駁的理論無法涵蓋到應該涵蓋的情景。無識人的可能性似乎沒有包含什麼雙方原本就同意接受的要點。所以無識人似乎對於物理理論難以構成威脅。

話是沒錯。沼澤人的故事還是有它的價值。當我們主張沼澤人沒有感質的時候，我們的想法是什麼？總不會僅僅只是一個可能性、僅僅只是在玩一個智力遊戲吧？我們能不能對這個可能性多說一點呢？

有一個想法是：一個個體有沒有心智可能跟演化有關。人類的心智是經由漫長的演化過程而出現的，沼澤人以及科學複製的李四則沒有演化史。所以沼澤人以及科學複製的李四都沒有心智，因為演化所需的時間絕對遠比單一個個體的生命要長很多。這個說法當然有一些怪異。照這說法，我們沒有一個人是有演化史的，因為有演化史的是人類，而非單一個個體。真的李四之所以有心智，是由於他是屬於有演化史的人類。沼澤人和複製的李四雖然在所有生理和物理層面跟真的李四完全一樣，他（它？）們跟真李四之間的差異只在於他們出



現的方式不同而已。我們為什麼不能承認他們跟任何真的人一樣，也有演化史呢？

另外一個想法是這樣的：或許個體有沒有心智必定涉及到個體（大腦）跟環境的交互作用，不論是意向狀態還是感質意識。缺少了這種與環境的交互作用就不會也不可能使得個體有心智。所以我們會認為沼澤人或科學複製的李四是沒有心智的。然而當我們繼續敘說沼澤人的故事，描述他開始如何覓食、如何避開危險、乃至於如何「學習」人的語言，很多人會開始覺得他是有心智的，或者他開始有心智了。他可能開始思考他周遭的環境，他可能開始預想下一步的行程，規劃明天的生活，他可能開始跟真的人溝通。

有些哲學家 and 認知科學家朝向考量演化與心智之間關聯的研究，其影響則是在二十幾年前新發展出來一門演化心理學。另外還有一些哲學家 and 認知科學家則愈來愈重視人與環境之間的互動。所謂的「意識行動理論」和「知覺行動理論」在這幾年也逐漸開展。我們期待這些新創的理論能提供我們更多對於心智的瞭解。



## 黑白瑪麗

### 反對物理論的理由之二

「黑白瑪麗」的故事是澳洲哲學家傑克遜提出來，用來反對物理論的一個思考實驗。他要我們假想下列的情景：

瑪麗從出生到現在，一直都被關在一間完全沒有色彩的屋子裡。這間屋子的奇特在於裡面的一切都是黑、白和灰色的。即使這間屋子裡的電視、電腦網路、手機……等，所有視訊設備的畫面也不例外，全部都沒有色彩。

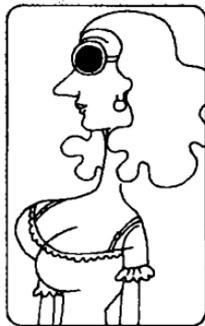
讓我們再進一步假想：科學家對於顏色以及色彩知覺的現象已經研究透徹，都已經窮盡，都已經有完整的瞭解，沒有任何遺漏。瑪麗具有極高的學習能力和超強的求知慾。她透過電腦、電視等視訊和通訊設備，學得了所有這些關於色彩以及色彩知覺的科學知識。不僅什麼東西是什麼顏色瑪麗知道得一清二楚，什麼顏色是什麼光波波長、人的大腦如何處理顏色的資訊進而產生色



彩知覺……等等，所有關於色彩和色彩知覺的物理知識、腦神經科學知識、乃至於其演化機制的知識，所有這些統稱為關於色彩的物理知識，不但科學家已經窮盡，而且瑪麗已經全部都學到了。

或許你還是有許多疑問。萬一瑪麗受了傷，她不就看到流出了鮮紅的血嗎？瑪麗的膚色也不是黑色或白色的。甚至，科學家指出，像瑪麗這種狀況，她的色彩知覺能力恐怕早就已經退化了，就算日後將她釋放出來，瑪麗也看不到任何顏色。為了祛除這些疑慮，我們可以對這情景做更嚴謹的設想。讓我們繼續假想上述狀況都被排除在外，釋放前的瑪麗是徹底活在看不到色彩的世界裡，但科學的發達足以確保她色彩知覺能力的正常。

現在，瑪麗終於被釋放出來了。在她眼前是一朵鮮紅的玫瑰、是綠油油的草地、是湛藍的天空……。瑪麗睜大了雙眼，興奮地說道：「原來紅色（例如紅玫瑰、紅番茄的顏色）看起來是這個樣子的。」釋放後的瑪麗終於知道原來紅色看起來是那個樣子，釋放後的瑪麗擁有了「紅色看起來是那個樣子」的知識。





這個「黑白瑪麗」的故事大致如此。哲學家想問什麼呢？敘說這個故事的哲學家傑克遜提出了他的問題：釋放後的瑪麗比起釋放前，有增加了任何新的知識嗎？傑克遜認為釋放後的瑪麗有增加新知識，也就是「紅色看起來是那個樣子」的知識。據此，他歸出一個結論：物理論是錯誤的！這個世界確實存在有非物理的事物。傑克遜藉著黑白瑪麗的故事所提出的，稱為知識論證，來論述物理論的錯誤。我將他的推想過程重新描述如下：

姑且讓我們假設物理論是對的。依據這個故事的敘述，科學家已經窮盡了所有關於色彩現象以及色彩知覺的知識，而且瑪麗在釋放前就已經學到了全部這些物理知識（包括光學、色彩知覺的腦神經科學、色彩的演化論……等）。換句話說，任何跟色彩和色彩知覺有關的物理現象都已經被人類發掘，都已經為瑪麗所知。所以，如果我們承認釋放後的瑪麗居然還有關於色彩的新知識，這項新知識絕對不會是關於色彩的物理知識。

由於釋放後的瑪麗確實獲得了新知識，所以這項新知識不是關於色彩的物理知識，這項新知識因而是關於非物理現象的知識，因此釋放後的瑪麗所知道的事情是非物理的。然而物理論是一



種承認只有物理世界存在的哲學主張，所以物理論是錯誤的。

釋放後的瑪麗所擁有的色彩知覺經驗是一種感質，所擁有的新知識其內容是「紅色看起來的那個樣子」。既然釋放後的瑪麗所獲得的新知識是關於這個感質的知識，這個感質的存在顯然不是釋放前的瑪麗所知道的。既然我們已經假設釋放前的瑪麗已經窮盡了關於色彩的一切物理知識，這個感質就不是物理的。這個感質的存在使得物理論為假。傑克遜進一步主張，感質不但是心理的（非物理的），而且是經由演化所產生的副產品；感質不具有因果力，但卻是由大腦活動所產生的。因為這個緣故，他的主張稱為副現象論。

我們要怎麼看待「黑白瑪麗」的故事呢？你接受傑克遜的「知識論證」嗎？物理論錯了嗎？黑白瑪麗的故事（亦即知識論證）在 1980 年代引發了許多的爭議。許多維護物理論的哲學家承認，這故事確實為物理論帶來極大的挑戰。他們甚至同意：如果接受知識論證的前提，確實會導致物理論的錯誤。因此，物理論哲學家試圖挑戰這個論證的前提。在黑白瑪麗的故事中，事先預設了釋放前的瑪麗能夠具有「完整的、關於色彩知覺的物理知識」。但是，有些物理論哲學家，如丘曲冷，就質疑為



傑克遜 (Frank Jackson) 澳洲國立大學哲學系教授，是澳洲幾位影響英美哲學界的重要人物之一。他的「黑白瑪麗」的故事可說是心與認知哲學領域人人皆知的。

什麼要接受這個預設。丘曲冷的說法是，目前物理科學還沒有進展到對於色彩知覺有徹底而完整的瞭解，也還沒有方法可以預估未來對於色彩知覺會發展出什麼物理科學，因此我們還沒有足夠理由接受黑白瑪麗故事中的預設。既是如此，即使承認瑪麗在釋放前就已經學到了關於色彩知覺

的全部物理知識，並不表示瑪麗在釋放後看見一朵鮮豔紅玫瑰時，就獲得了新知識。

我並不認為丘曲冷的說法真的回應了知識論證，因為他並沒有真的排除了「黑白瑪麗」的可能性。「黑白瑪麗」的故事不是一個純粹的邏輯可能而已。丘曲冷不能以這個可能性沒有足夠依據就加以解消。多數物理論哲學家選擇正面迎戰知識論證。他們試圖指出：即使接受這個預設，知識論證是否真能駁斥物理論，仍然有待商榷。這裡，物理論哲學家分進合擊，分別採取了下列兩種策略：能力假設以及表徵假設。

主張「能力假設」的哲學家主要是澳洲的路易斯。他跟丘曲冷不同，他承認瑪麗釋放後確實學到了新東西，



不過瑪麗所學到的是新的能力，而不是新的知識。什麼能力呢？一種辨別紅色跟其他顏色的能力，以及再辨認紅色的能力。這是路易斯所說的能力假設。按照能力假設，知識論證自然不足以否定物理論。因為物理論的主張是：這世界所存在的僅僅只是物理的，並沒有非物理的事物。對於物理論的否定所需要的前提，必須是能夠讓我們承認非物理世界的事實的。因此，黑白瑪麗的故事必須能夠做到，釋放後的瑪麗在看到紅玫瑰的時候，所獲得的知識是關於非物理的事實。如果釋放後的瑪麗所獲得的只不過是辨認與再辨認紅色的能力，這個故事是沒有辦法駁斥物理論的。

我想不可否認的，在黑白瑪麗的故事裡，我們應該承認釋放後的瑪麗確實比起釋放前的她，多出了辨認與再辨認紅色的能力，這應該是無庸置疑的。可是這並不足以表示她就沒有獲得新的知識。讓我們再回頭想一想：如果我們要承認釋放後的瑪麗擁有新的知識，我們會如何描述這個知識呢？瑪麗當然可以對於她所辨認以及再辨認的事物有一些想法。而瑪麗的想法就是：「原來所謂



在當代知識論裡，「知識」至少包括兩個類型：「能力知識」，例如知道怎麼游泳，以及「命題知識」，例如知道游泳池的位置。這些物理論哲學家有時候會說，瑪麗是獲得新知識，但其實指的是能力知識，而這裡所關心的主題僅僅是命題知識而已。



「看到紅色的知覺經驗」，就是我現在看到玫瑰的顏色之後，所產生的這個知覺經驗。」我認為能力假設並沒有替物理論回應到黑白瑪麗的挑戰。

到此為止，物理論哲學家顯然必須要指出：釋放後的瑪麗所獲得的即使是新的命題知識，仍然不會否定了物理論。怎麼做呢？所謂的「表徵假設」就是從這地方著手。這個假設也是丘曲冷提出來的。我們知道同一件事實可以用不同的方式呈現出來。從認知主體的角度來說，認知主體對於同一件事物可以用不同的方式加以表徵。我們提到的「孔明與諸葛亮」的例子以及伊底帕斯的例子都說明了這點。這帶給了哲學家新的想法。釋放前的瑪麗確實已經窮盡了關於色彩的所有物理知識。釋放後的瑪麗確實獲得了新知識。但其實那是同一件物理事實以不同方式呈現給瑪麗而已，就如同孔明可以用「劉備最重要的軍師」來呈現，也可以用「設下空城計的那個人」來呈現。同樣地，釋放前的瑪麗是以物理學、生理學、腦神經科學的方式來表徵「紅色知覺經驗」這種感質的；釋放後的瑪麗則是以親身經歷的方式來知道「紅色知覺經驗」這種感質的。所以在提出表徵假設的哲學家看來，黑白瑪麗只是對於相同的物理事實有不同的表徵而已。

按照表徵理論的回應，物理論確實沒有被黑白瑪麗



的故事擊倒。副現象論又要如何回應呢？或許我們還要再細想表徵理論的說法。黑白瑪麗用物理科學來表徵「紅色知覺經驗」這種感質跟他之親身經歷「紅色知覺經驗」這種感質，真的是對於同一事物的不同表徵嗎？我們曾經強調過，感質這種心理狀態有一種質性的特徵。這種質性的特徵似乎不是以物理學的方式來表徵感質這種心理狀態時，所能掌握到的。或許我們應該從這裡下手，繼續思考感質的問題。

## 身為蝙蝠的樣子

### 意識究竟是什麼？

我們知道蝙蝠是會飛的哺乳類動物，但是蝙蝠沒有視覺，它是透過它獨有的迴音偵測系統來行動的。假設蝙蝠跟人一樣是有意識的，哲學家納格問：我們有可能理解它的意識嗎？科學有可能研究它的意識嗎？當然，討論蝙蝠只是作為一個例子而已。納格舉蝙蝠作例子，是因為蝙蝠這種哺乳類跟人類非常不同，但可能是有意識的。如果你對於蝙蝠有無意識產生疑慮，我們可以找其他的例子。



隨著科學的發達，我們對於蝙蝠的瞭解愈來愈多，牠的生理構造，牠的迴音偵測系統的運作……等，科學家不斷累積研究的成果。那麼，蝙蝠以迴音偵測系統飛行、偵測獵物等，牠內心所經歷的，能夠被科學研究出來嗎？我們能夠知道牠那個樣子的意識經驗嗎？納格指出，我們不可能知道蝙蝠的意識經驗，因為蝙蝠的意識經驗具有那個樣子的特徵。若是依照納格的意思，那麼「身為蝙蝠是什麼樣子？」是一個科學上無解的問題囉？

納格 (T. Nagel) 紐約大學哲學系教授。他所提的「蝙蝠」故事在心與認知哲學界可說「家喻戶曉」，對於科學如何研究人類心智提出了重大的挑戰，是認知科學研究「意識」問題時絕對不可能輕忽的。



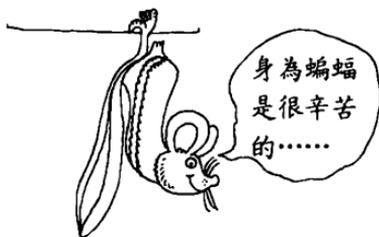
為什麼呢？納格的說法是意識經驗具有觀點性，而且是第一人稱的觀點，不是第三人稱的（各位不必考慮「第二人稱」，因為我們不是在談文法，而是在區別「我」以及「我以外的任何人」，也就是「他人」）。第一人稱的觀點就是自己的觀點，第三人稱的觀點就是他人可以擁有的觀點，也就是平常所說的「客觀」。這些概念在哲學說來有些複雜。讓我用比較合乎一般人直覺的方式來說（請讀者回顧我在第二章解說的「客觀—主觀」的幾



種意義)。

有「觀點」當然就有被觀的對象，所以我們從被觀的對象來思考。一塊礦石可以被「觀」——被觀察、被研究、被瞭解、被知道。對於礦石的瞭解並不限定只有從我自己的觀點才能掌握，簡單說，任何人都能瞭解礦石的性質，知道礦石是什麼。不僅礦石是如此，所有物理世界的物體、性質、現象都是如此。這是為什麼哲學家通常將物理世界稱為客觀世界的緣故，這也是為什麼有哲學家用「第三人稱的存在」來描述物理世界。

但是許多哲學家認為人的意識經驗跟物理世界非常不一樣，意識經驗是一種第一人稱的存在。納格舉蝙蝠作例子就是試圖描述意識經驗這種第一人稱的特徵。意識經驗必定是以「那個樣子」來呈現給那個認知主體的。我吃柳丁所經歷的那個樣子的酸澀感是我的觀點下才能體會的，你吃柳丁所經歷的那個樣子的酸澀感是你的觀點下才能體會的。寒天飲冰水，冷暖自知！





由於意識經驗必定具有這種第一人稱的觀點性，但是科學是一種第三人稱觀點的研究，所以沒有任何科學可以研究出意識經驗的本質，沒有任何人可以知道蝙蝠那樣子的意識經驗，除了蝙蝠自己。同樣地，如果必定要從我的觀點才能體會我吃柳丁所經歷的那個樣子的酸澀感，如果必定要從你的觀點才能體會你吃柳丁所經歷的那個樣子的酸澀感，科學如何從我的觀點來研究我的那種酸澀感，科學又如何能從你的觀點來研究你的那個樣子的酸澀感？

也許我可以想像我是一隻蝙蝠。我想像我有蝙蝠的膜翼，有蝙蝠的迴音偵測系統；我想像自己藉著迴音偵測系統像蝙蝠一樣飛行，想像自己經歷到蝙蝠所經歷的意識經驗。不過，我真的想像到了蝙蝠的意識經驗了嗎？如果我要瞭解你吃柳丁所經歷的那個樣子的酸澀感，也許我可以想像我是你，或者乾脆我自己吃一顆柳丁。即使如此，我自己吃柳丁時所經歷的那個樣子的酸澀感仍然不會是你吃柳丁時所經歷的那個樣子的酸澀感。雖然同樣都是酸澀感，你那個樣子的酸澀感是從你的觀點才能體會的。

其實想像力跟這裡的哲學論旨一點關係也沒有。不論我們的想像力是薄弱還是豐富，當我想像自己是蝙蝠的時候，我還是帶著我自己的觀點去想像的。蝙蝠的意



識經驗是蝙蝠從牠自己的觀點來經歷的，跟我如何想像無關，因為我的任何想像都是帶著我的觀點來想像，帶著我自己的觀點來經歷那意識經驗，終究還是我自己在經歷那意識經驗。

意識經驗由於必定具有觀點性，而物理世界不具有觀點性，不僅使得心理世界跟物理世界有很大的差別，也同時使得我們面臨科學研究意識的一大障礙。那麼意識經驗的觀點性哪裡來的呢？

我提過物理世界不具有主體性，心理世界（主要是意識）具有主體性。主體性是區別「心」與「物」的根本特徵之一。讓我們回顧我在第二章關於「主體性」的解說。我們知道所謂意識經驗具有主體性，意思是說，如果有個意識經驗出現，則那個意識經驗必定存在於嚴格意義下的擁有者。我的意識經驗只有我能夠擁有，我是我意識經驗嚴格意義下的擁有者，也就是說，我的意識經驗不可能改變擁有者，變成別人的意識經驗。意識經驗的存在，跟它的擁有者之間有必然的聯結，它們無法被轉移。我們知道現代醫學發達，已經能夠進行器官移植，將某人身上的腎或肝移植到別人身上。可是科學能夠做到將某人高興的感覺移植到別人身上嗎？這似乎是做不到的！我的高興就是我自己才能體驗到，你怎麼可能體驗到我的高興的感覺呢？哲學家因此說，意識經



驗具有主體性。

意識經驗的觀點性其實就是意識經驗的主體性。既然我的意識經驗必定以我作為擁有者，對於我的意識經驗的瞭解，當然只有從我自己的觀點才作得到了。既然你的意識經驗必定以你作為擁有者，對於你的意識經驗的瞭解，當然只有從你自己的觀點才作得到了。

許多哲學家認為，如果承認意識經驗必定具有這種主體性和觀點性，物理論就錯了。因為物理論只承認物理世界的存在，而物理世界是第三人稱的存在，意識經驗是第一人稱的存在，所以意識經驗不可能屬於物理世界，不可能等同於物理世界的任何性質或現象，也不可能被研究物理世界的科學所掌握。有些哲學家和科學家因而走向了二元論的哲學立場。

那麼意識究竟是什麼呢？這就是「主體性」為研究心與認知的哲學家 and 科學家所帶來的最大的困惑！意識「主體性」問題是如此的棘手，以至於吸引了全世界一流的人才，在努力解開它的謎團。或許十年後我們會知道意識是什麼，或許我們還要等上一百年。究竟能不能找出答案，我的讀者，請你也一起動動腦吧！



牠!



或他!



## 會算術的馬

很多養過寵物的人常常覺得他的寵物跟他很有默契、心有靈犀；尤其是養狗的人更覺得狗是明顯有喜怒哀樂的動物，牠們會跟主人親熱，會叼鞋子給主人穿、會看家。我們甚至還能訓練出導盲狗、救難犬等，來幫忙我們。愛因斯坦所說的「訓練有素的狗」對專家來說，或許是種貶抑，但對狗絕對是種讚賞！牠們會思考、有判斷力，這既叫人驚訝、卻又習以為常！動物是不是有心智對大多數人並不是個問題。牠們當然有——至少比較高等的動物一定有。



對我漢斯  
來講，都沒差！

這些都還不夠神奇。在 20 世紀初，德國有位馬戲團的馴馬師養了一匹名叫漢斯的馬。這位馴馬師發現漢斯竟然有計數的能力，而且還會作加減的計算。不但如此，漢斯能夠拼出人名（用德文）。牠當然不是用說的，牠是用輕輕點頭或者輕輕踏前蹄的方式來給答案的。例如，在拼字的時候，牠會輕輕踏前蹄一次代表字母 A，輕輕踏前蹄兩次代表字母 B；依此類推。而且牠會在字母跟字母之間間歇一下。（漢斯大概沒有拼過用到 XYZ 等字母的字吧！）「聰明漢斯」不



僅成了主人的搖錢樹，牠的聲名遠播，儼然成了當代動物界的傳奇（靈犬萊西是另一個）。

話雖如此，漢斯究竟有沒有那麼聰明，不無疑問。在當時有幾位科學家風聞漢斯驚人的表現，來到了馬戲團，作了一連串的觀察和實驗，希望能找出一些解釋。他們發現漢斯的表現是真實的，不是刻意操縱的。不過這些科學家也發現，當主人不在現場時，漢斯的表現就差很多。可是他們也確定主人並沒有作假。那麼到底是怎麼一回事呢？這些科學家發現，事實的真相是：漢斯並不知道正確答案是什麼，但是當牠輕輕踏著前蹄的次數是正確的時候，牠能夠感受到主人以及觀眾的反應，而停住牠輕輕踏前蹄的動作。事實上，漢斯並不知道答案，牠藉著出現在人類（主人及觀眾）肢體語言上的變化，來給出相對的反應而已。牠的這種感受相當靈敏，但是牠不知道如何計數，如何加減。知道答案的是人類，然後反映在人類的肢體語言，然後被漢斯感受到，最後漢斯做出人類以為

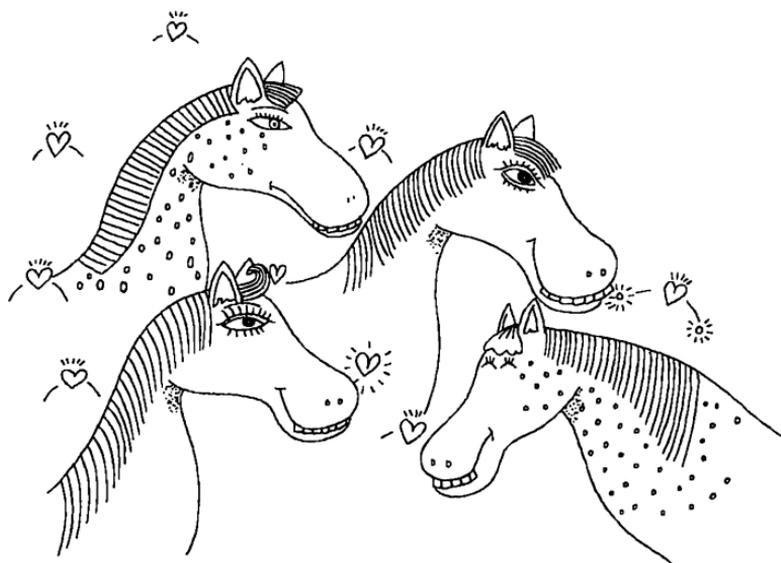




給出了正確答案的適當行為。用行為主義的話來說，牠被制約了。漢斯的行為只不過是一面鏡子，人類從這面鏡子看到自己給的答案。

漢斯算是相當聰明了，但牠還沒有聰明到能夠作出算術、拼出人名的程度。那麼動物有心智嗎？這裡提醒讀者，漢斯既然是一匹活生生的馬，當然會感受牠週遭的事物。哲學家關心的是動物是否有心智能力，以及是否有與人類比較相近的意向性狀態，例如思想、信念等。

在哲學史上笛卡爾主張動物是沒有心智的。牠們的



呼呼!

還有什麼比現在 1...2...3... (個) 更幸福呢!



一切看似有心的行為，都可以用機械的方式加以說明。對於動物任何行為表現的解釋，其實跟解釋鐘錶運作的原理相差不遠。笛卡爾說，動物只不過是比較複雜的機器而已。「聰明漢斯」的故事就是最好的例子。我們人類只是將自己的內心世界、自己的思維投射到動物身上，用我們理解人類行為的方式來理解動物行為而已。有養寵物的人只不過是將自己的思維投射到自己的寵物身上，制約了寵物的行為，以至於誤以為寵物是有靈性的。

笛卡爾曾經提出兩個測試來作為動物，乃至於任何個體，是不是有心智的判準：語言測試以及行動測試。任何不能通過這兩個測試的個體，就不是有心智的。笛卡爾之所以主張動物沒有心智，就是因為他認為動物不能通過這兩個測試：一方面動物沒有辦法使用文字表達想法；另一方面動物沒有理性，不能用思想來決定要如何行為。

然而究竟動物能不能通過笛卡爾的兩個測試呢？究竟這兩個測試是不是好的、可以用來判斷任何個體有沒有心的判準？

達爾文跟笛卡爾的立場恰恰相反。他認為有一些動物有比較低階的心智。跟生物的其他特徵一樣，心智都是經歷長久時間而演化出來的。人類固然演化出了心智，大自然並不排除人以外的動物也能演化出心。每個物種



可能發展出最適合自身需要的心，其差別是程度上的，不是種類上的。當代新興了一門學問，稱為動物認知行為學。其中有一些科學家也提出類似的說法，狗有狗的心智，猴子有猴子的心智，每一個生物都依照他們存活在世界上的需要來發展，在同一個動物種族之中也有較聰明或是較適應環境的個體。

當代哲學家卓斯基跟笛卡爾的立場也相反，他跟達爾文同樣，認為動物是有心智、有思想的，但他不是從演化角度來主張動物有心智，而是認為動物之所以有心智，是由於動物有學習的能力。他舉了一個例子：



達爾文 (Charles Robert Darwin, 1809-1882) 英國生物學家，進化論的創立者。1859年發表《物種源始》，提出生物進化的原則是物競天擇、適者生存。此一理論對生物學的發展造成極大的影響。

卓斯基 (Fred Dretske) 史坦福大學教授。近年退休後，轉往杜克大學服務。他首先將資訊理論引進對於知識論的探討，並在心與認知的哲學領域裡，對於「心理表徵」提出「資訊語意學」的主張，近年更進一步將他所發展的表徵理論應用在對於「知覺」的說明。是心與認知哲學領域中非常重要的哲學家。





一隻覓食中的鳥吃掉一隻塗上有毒物質的甲類蝴蝶，並且因此生病感到難受的噁心。下一次當牠看到這類蝴蝶時，牠便不會再獵食，甚至當另一種跟這類蝴蝶很類似的乙類蝴蝶出現在牠面前時，牠也不會獵食。

卓斯基認為這表示這隻鳥「相信」那一種外形、那一種氣味的蝴蝶會使牠生病噁心，這隻鳥會這樣「想」，而這個「想法」使得牠即使見到那些類的蝴蝶也不會採取獵食的動作。再把實驗擴充：

假使有一天這隻鳥非常非常饑餓，牠迫不得已再度嘗試獵食一些氣味和外形跟原先的甲類蝴蝶相像的蝴蝶。如果這一次牠獵食的是乙類蝴蝶，或者是沒有塗上有毒物質的甲類蝴蝶，那麼牠很可能會再度開始獵食類似甲類蝴蝶的蝴蝶。

卓斯基認為這隻鳥有學習的能力、有記憶、也有信念。牠是有心智的。

認知動物行為學家對於動物作了許多觀察和實驗，以下是可以用來承認與心智有關的一些動物行為的研究主題：欺騙、遊戲、概念、「自我」概念、溝通、作標記



等。這些主題都相當有趣並且值得深思，可惜本書無法一一檢視動物認知行為學的研究成果，我將只挑其中一些來介紹。

## 如何問「動物有沒有心智」的問題？

認知動物行為學要如何進行動物有沒有心智的研究呢？由於我們之前曾經將心理狀態分為感質和意向狀態兩類，所以要回答這個問題，就要考慮動物會不會擁有這些心理狀態。但是如果要进行研究，認知動物行為學還是得從動物行為下手。就這一點來說，動物心智的研究問題跟他心問題是非常類似的。

拓夫茨大學哲學系教授丹尼特曾經將「心」作了四個等級的劃分：達爾文式的心、史金納式的心、巴柏式的心、以及葛雷哥式的心。「達爾文式的心」指的是能展現生物本能的行為。「史金納式的心」是指個體的行為能夠經由制約的手段而建立出來。「巴柏式的心」是指能夠純粹靠思考來預估可能發生的情況，藉以調整個體自己的行為。最後，「葛雷哥式的心」是指具有自我意識以及自我概念，能夠對自己的經驗與想法進行反省。

我認為丹尼特的說法只是在混淆「心」的概念而已。照他的說法，所有生物都有心智，因為所有生物都有達



爾文式的心，也就是對於環境刺激能做出生物本能的行為反應。我並不是反對「心」的概念不能這樣使用，而是說這樣的用法對於我們想弄清楚的「動物有沒有心智」的疑問一點幫助都沒有，他的用法也使得我們之前提到的心物難題，以及取消唯物論和各種物理論的爭議都變得一點也不有趣。

儘管如此，丹尼特的說法反倒提醒我們：「行為」可以分為生物本能的、制約的、以及意向的（在第四章我們曾經提到這個區別）。如果科學家要從動物行為來研究動物的心智，科學家必須找到屬於意向行為的類型，才能決定動物是不是有心智。

話雖如此，仔細想想，「動物有心智嗎？」這個問題有兩重困難。首先，「動物」一詞涵蓋的範圍實在太廣了。從天上飛的老鷹、鴿子和麻雀，到地上跑的老虎、獅子和羚羊，到在水裡游的鯨、鯊魚、蝦、蟹、和八爪魚；從沙漠裡的蜥蜴、蠍子，到叢林裡的猩猩、豪豬，到雨林的鸚鵡、大蟒，全部都是動物。至少從表面來看，主張猩猩有心智比起主張蝦子有心智要讓人容易相信得多了；主張海豚有心智比起主張蝴蝶有心智要讓人容易相信得多了。問「動物有心智嗎？」這種一竿子打盡的問題，不論要回答「有」還是「沒有」似乎都有些道理，也似乎都少了些道理。



另一方面，問動物是不是有「心智」是太籠統的問法。我們說過，對於「心智」的瞭解一方面包括各種認知機制，另一方面包括各種心理狀態，而後者又包括「感質」和「意向狀態」兩大類。姑且擱置認知機制不談，感質這一類所涵蓋的心理狀態本身就有很大的差異。痛覺和癢覺跟我們的身體關聯比較密切，歡欣與嫉妒跟我們的身體就沒有那麼大的關聯了。感官知覺經驗來自於五官的運作，心情和感受跟五官的運作不太有關。意向狀態這一類大概也是相當龐雜。判斷和希望有很大的差異，想像和相信非常不同，自我反省跟揣測他人更是非常不一樣。對於「動物是不是有心智」這個問題的探討，不同的物種展現的心理狀態與心理能力可能是不一樣的。

令人遺憾的是，正由於「動物是不是有心智」的問題有這兩重困難，所以我們會看到主張動物有心智的人跟反對動物有心智的人，雙方未必是在爭辯相同的事情。笛卡爾在否認動物有心智的時候，他並沒有否認動物有感覺、情緒等感質，他否認的是動物具有思維——別忘了，笛卡爾是以「思維」作為心智的本質。

我們可以歸納出幾種關於動物心智的哲學立場：

1. 任何動物都沒有心智。
2. 有些動物有心智，其他動物則沒有；而且在那些有心



智的動物之中，彼此展現的心理狀態與認知能力未必都是一樣的。例如有些動物有欺騙的行為，有些動物則沒有。至於哪些類的動物有心智，以及不同類的動物擁有的是什麼方面的心智，則需要認知科學作進一步的研究。

3. 所有動物都有心智，但是各種動物展現的心理狀態與心理能力未必都是一樣的。同樣地，認知科學對於動物心智的研究是不可或缺的。

由於「動物是不是有心智」這問題有上述雙重困難，所以我們不可能找到一個標準，來作為動物有沒有心智的依據。相反地，我們應該就各種主題，例如有沒有概念、有沒有思想、會不會欺騙、有沒有發現被騙的能力、有沒有反省能力、有沒有自我概念、有沒有「他」的概念、社會概念、有沒有罪惡感、有沒有羞恥感、有沒有懲罰行為……等，來思考各種動物有沒有心智的問題。

接下來我將介紹認知動物行為學的一些研究，以分別考量幾種動物是不是有心智的可能性。不過在這之前，讓我先轉向另外一個密切相關的議題：擬人論。

## 擬人化的思維

1. 「擬人論」簡單說，是一種將人以外的動物（甚至



於植物、乃至於非生物，如礦石、星球……等）當作是人的一種作法。擬人化的作法最常出現在神話、童話、科幻小說、詩詞文學等，但最為人知的莫過於卡通了。如果我們以及科學家在說動物有心智時，我們只不過是採用了擬人的說法，我們並不是承認動物有心智；但是如果我們認為對於動物行為的瞭解不是擬人化，而是實實在在的理解，我們就是承認動物有心智的。

不過擬人化的作法需要進一步的說明。根據費雪的主張，擬人論可以依照用法區分為虛構擬人論以及詮釋擬人論。虛構擬人論出現在童話故事、神話、卡通等，透過創作者豐富的想像力，這些故事裡的動物、植物、甚至石頭、太陽都會說話、會有喜怒哀樂、會有慾望、猜忌、會思考、判斷、推想。與其說卡通和童話裡的石頭、植物和動物有人一樣的行為和語言，倒不如說這些角色都是人類以石頭、植物和動物的外形出現罷了！這種屬於創作型的擬人論在文學藝術娛樂方面有價值，在學術研究上就沒有價值了，因為這種作法不可能具備解釋現象的能力。

學術研究，尤其動物認知的研究，多半採用詮釋擬人論的作法：將動物類比成人類，以解釋或理解動物的行為。例如將幼狼之間的活動當作是在玩耍，將黑猩猩露出牙齒的動作當作是發怒等。對於動物心智的研究顯



然讓我們想起了對於他心問題的探討。心理因果原則以及類比法同樣是動物研究所必要的。由於我們已經看過了它們的原理以及缺點，所以這裡就不再重複（見第三章）。

由於詮釋的擬人類比法使得對於動物行為的解釋容易進行，因此受到相當的歡迎。不過，究竟對於動物行為的瞭解只不過是一種詮釋的擬人化而已，還是動物真的具有引起那些行為背後的心理原因呢？有些哲學家有不同的意見。

哲學家對於擬人論最大的批評在於擬人化的作法犯了範疇謬誤。人跟動物本來就是不同的，把描述人類特性的詞語拿來描述動物，基本上就不適當。大多數反對擬人論的觀點都蘊涵對於動物心智的否認。他們認為將意向性加入我們對於動物行為的理解和詮釋，並假設動物具有情感、理性、意圖、判斷，這些只不過是我們人類作為觀察者的時候，所做的一種內心投射，並沒有科學的依據。動物的行為很可能只是本能反應或者環境刺激的連結而已，沒有我們所認為的種種內在心理因素。例如，孔雀開屏不是王者風範，而是向雌孔雀提出交配的訊號。難道我們要承認賽馬場中極力奔跑的馬心中想著要跑贏這場比賽嗎？難道我們真的認為小鳥的叫聲是在唱歌、在表達牠內心的高興嗎？



有趣的是，大多數動物學家則極力反駁哲學家所指責的範疇謬誤。他們認為這是人類的傲慢，自以為只有人類才有心智。有什麼理由認為人跟動物的不同就是有心智跟沒有心智的不同？以擬人論來對動物進行研究，為什麼就犯了範疇謬誤了呢？提出範疇謬誤這種指責的人反而阻礙了對於動物做科學研究的嘗試。

我想反對「範疇謬誤」這種指責的人是有一些道理的，但也同時需要一些修正。確實大多數認知動物行為學家使用擬人法來研究動物行為，也確實大多數認知動物行為學家主張動物有心智。不過純粹從研究方法的角度來看，擬人論是中立的，既不因此表示使用擬人論作研究的科學家一定承認動物有心智，也不表示他們一定否認動物有心智。動物是不是有心智的問題跟擬人論的使用是互不衝突的；擬人論跟上一節所說關於動物心智的三種哲學立場都相容。使用擬人論來研究動物的行為，目的在於找出被研究的動物，在暫時接受牠們有心智的假設下，會呈現哪些心理現象和心理能力。

其實我對於以擬人論來研究動物行為深深質疑。這裡我簡單作兩點觀察：首先，要將動物行為類比到人類行為通常不是一件容易的事。絕大多數的動物外形跟人類差異非常非常大，例如鳥類、爬蟲類、和魚類……等。我們是在什麼意義下，將動物行為類比到人類行為呢？



如何類比呢？牠們沒有四肢，不會像人一樣走路、跑步、用手取物……等。我們也不會像鳥類一樣有拍翅膀的行為、啄羽毛的行為；像魚類一樣在水裡用鰓呼吸的行為、擺動尾巴游泳的行為……等。牠們和人類之間有多少行為是比較能拿來作類比的呢？

其次，動物的知覺系統跟人類的非常不一樣。狗能聽到一些人類聽不到的音波頻率，這是我們很熟悉的事了。此外，還有很多動物有不同的知覺系統，比如蝙蝠的迴音偵測系統，生存於地下的動物所具有的震動感應，沙漠地帶的蛇蠍類的熱感應……等。這些都是人類沒有的知覺系統，所以我們人類也不會出現經由這些知覺系統而產生的行為。我們如何對這些行為進行類比呢？我們如何進行擬人化呢？

第三，從心理世界產生的行為往往是非常複雜的；有些心理狀態是很難從行為去推斷的。例如，罪惡感、羞恥感、發現被騙了、報復、報恩……等。說實在的，我們不是很明確這些心理狀態會產生什麼樣的行為，我們很難確定要從什麼行為來推斷牠們是不是處於這類心理狀態。

我並不是主張擬人論犯了範疇謬誤，而是指出擬人論在動物行為研究上有一些實質的困難。對於動物是否有心智的問題，也未必能做出最好的解釋。



## 動物的語言與概念

很多人都知道「蜂舞」：蜜蜂可以透過身體的擺動、以轉圈的方式告知同伴花蜜所在的地點。這個行為算不算是蜜蜂的語言？如果語言的功能是溝通，那麼蜜蜂的舞蹈的確達到了溝通的目的。由於溝通是一種社會行為，蜜蜂必須形成社會組織，並且以同樣的行為模式交互作用，溝通才有可能。如果蜜蜂出現溝通的現象或許可以承認蜜蜂是有心智的。

不但如此，甚至還有科學家認為蜜蜂有思考的能力，因為蜜蜂在尋找花蜜帶回巢穴，告訴其他蜜蜂花蜜的來源，以及如何到達來源之間，展現了重組視覺資訊的能力、聯想回憶的能力、分類視覺資訊的能力、以及整理資訊前後關聯性的能力。

為了更進一步瞭解蜜蜂，科學家作了一些實驗，試圖訓練蜜蜂能辨認顏色和圖案。這實驗是這樣的：設計一個 Y 字形的迷宮，其入口為 Y 字形下方頂端，並且在入口處會標示藍色或者黃色，而在 Y 字形的交叉點，對於兩個岔路一個標示藍色，另一個標示黃色。實驗的目的是蜜蜂必須選擇和入口處一樣的顏色，才能得到糖蜜作為獎賞。令人訝異的是，實驗結果顯示，蜜蜂確實能



夠正確區別差異。因此科學家結論認為蜜蜂是會思考的。許多科學家認為同樣的結論可以套用在所有表現類似行為模式的動物，藉以判斷動物是不是具備思考的能力。

有位科學家花了幾年的時間教導一隻非洲灰鸚學習「相同」和「相異」兩個概念，以及物體的數量、顏色、形狀、組成份。他讓這隻鸚鵡面對三個黃色的塑膠製品、一個紫色的塑膠製品、一個紅色的塑膠製品、以及一個綠色的木頭製品。然後問這隻鸚鵡一些問題。經過一段時間的訓練之後，這位科學家發現，如果問這隻鸚鵡：「哪一個材質是綠色的？」牠會回答木頭。而問牠有幾個是黃色，牠也可以給出正確的答案。

科學家研究非洲的一種長尾黑顎猴，發現了一些很有趣的現象。長尾黑顎猴能針對不同的天敵發出不同的警示聲，並做出相應的避敵動作。當牠們發現老鷹的時候，牠們發出某種警示聲通知其他同類，然後躲進矮樹叢裡；當牠們發現黑豹，會發出另一種警示聲以通知其他同類，然後逃到樹上；當牠們發現蟒蛇，牠們又會發出另外一種警示聲，然後快跑以避開蟒蛇的攻擊；而當牠們發現人類在附近的時，牠們再度發出另一種不同的警示聲以通知同類，然後不動聲色，靜靜伏在樹叢裡。這些警示聲明顯是牠們的溝通方式。更有趣的是，如果族群裡有兩隻長尾黑顎猴在爭吵，處於劣勢的那一隻會



爬到樹上發出黑豹警示聲——即使當時並沒有黑豹的跡象，然後所有長尾黑顎猴都會迅速爬到樹上，而結束這場紛爭。

電視 Discovery 頻道曾經報導過，生活在非洲濃密叢林裡的一種黑猩猩。這種黑猩猩是群居的，當牠們移動的時候，會使用大片的樹葉來留下記號，指引同伴。黑猩猩可以透過別的同類留下的記號，知道其他同類做了些什麼，甚至有科學家就利用這種記號找到了藏匿中的黑猩猩群。當不同猩猩群的移動路線交錯，以至於造成記號交錯的時候，黑猩猩群的領導者會重重的踐踏附近的植物，或是在地面上放置大片的葉子來指引黑猩猩群移動的方向。

蜂舞是語言嗎？長尾黑顎猴的警示聲是語言嗎？如果純粹就溝通來看，牠們的行為（包括聲音）確實是一種溝通的形式，也達到傳遞訊息的目的。不過，要將蜂舞和警示聲當作是一種語言，恐怕言之過早。當然這必定涉及我們如何瞭解「語言」。就人類的語言來看，語言不僅有語法（亦即文法），還有語意。語法決定哪些表達方式是可被接受的，哪些則是不被接受的。例如「玫瑰是紅色的」是可被接受的中文表達方式（不管是用寫的，還是用說的），但是「色玫紅是瑰的」雖然用的都是中文字，並不是中文的適當表達方式。語意是另外一種決定



表達方式能不能被接受的原則。例如「玫瑰是空白的」雖然語法上沒有問題，卻沒有表達任何意義，所以是不被接受的中文句子。

語言之有意義在於它表達了思想，思想的一項重要特徵是我們在第六章提過的「語意組構性」。因此，蜂舞和長尾黑顎猴的警示聲是不是語言，就要看它們有沒有具備這種語意組構性。然而蜂舞和長尾黑顎猴的警示聲並不具備這種語意組構性。所以從這點來看，它們不是語言。不但如此，黑猩猩之用樹葉留下記號來跟其他同類聯絡，也算不上是一種語言。

蜜蜂和長尾黑顎猴既然沒有語言，牠們有思想嗎？這又涉及到語言跟思想之間的關聯了。如果擁有思想是擁有語言的必要條件，亦即沒有思想就沒有語言，那麼動物之沒有語言並不表示牠們就沒有思想。相反地，如果擁有語言是擁有思想的必要條件，亦即沒有語言就沒有思想，那麼動物之沒有語言當然表示牠們就沒有思想。語言和思想之間究竟具有充分條件的關係、必要條件的關係、還是兩者都有？

讓我們轉向另外一個層面來思考動物有沒有思想的問題。

我們曾經遭遇過「指涉隱蔽難題」，指涉隱蔽的現象是思想的一大特徵。從這點來看，蜜蜂的蜂舞和長尾黑



充分條件與必要條件：

- ◆ 當我們說：「A 是 B 的充分條件」時，意思是「如果 A 為真，則 B 為真」。例如，「張三是單身漢」這句話如果為真，則「張三是男的」這句話就為真。此時，我們說第一句話所說的事情是第二句話所說事情的充分條件。
- ◆ 當我們說：「A 是 B 的必要條件」時，意思是「如果 B 為假，則 A 為假」。例如，如果「王五是男的」這句話為假，則「王五是單身漢」這句話為假。我們說前一件事情是後一件事情的必要條件。

顎猴的警示聲恐怕沒有表達思想，因為蜂舞和警示聲並沒有表現出指涉隱蔽的現象。黑猩猩用樹葉留下記號的行為也同樣沒有表現出指涉隱蔽的現象，所以也算不得是牠們有思想的證據。

其實我們還可以從前面提過的行為分類的角度，來審視這些動物行為的觀察，包括蜜蜂迷宮的實驗以及鸚鵡的故事。「聰明漢斯」的故事已經告訴我們不要輕易被表象所蒙蔽。在蜜蜂迷宮的實驗以及鸚鵡故事裡的動物，牠們所產生的行為會不會只是制約的結果？動物的行為究竟是受到思想的引導、或者僅僅是本能的反應、或者是制約的結果，都有待進一步的研討。我們不能太過輕率下判斷。



我建議一個簡單的策略作為參考：如果能夠使用「制約」概念來理解動物行為，就不必使用「意向」概念來理解動物行為；如果能夠使用「生物本能反應」的概念來理解動物行為，就不必使用「制約」概念來理解動物行為。按照這個策略，我認為蜜蜂迷宮的實驗以及鸚鵡的故事所反映的是：牠們的行為是被制約的，而不是牠們真的有思想。同樣地，蜜蜂的蜂舞、長尾黑顎猴的警示聲、以及黑猩猩用樹葉留下記號的行為可以用生物演化的生物反應來說明——儘管這種生物反應在演化適應性的表現上，比起含羞草一被碰觸葉子就收縮，更複雜得多。

這個策略在實施上當然還是非常粗糙。畢竟在我們對於「制約機制」以及生物演化的「適應行為」有更充實的瞭解之前，我們很難斷言被研究的那項動物行為究竟是純粹生物性的？是制約的結果？還是內心活動所造成的？無論如何，讓我在下一節再討論一個有名的動物研究，作為剛剛所說研究策略的一個例子，以作為參考。

## 欺 騙

欺騙行為在人類社會生活中屢見不鮮，如何發現欺騙卻不是一件容易的事。測謊機的原理是假定人類在說



謊的時候，會產生膚電的變化或者其他身體的生理變化。然而，即使測謊機是高度可靠了，欺騙不一定來自於說謊的行為，當欺騙發生時，也不太有機會用測謊的方式來發現（除非你能逮捕可能的欺騙者）。對於動物的欺騙行為似乎就更難發現了。時至今日，我們唯一能夠用以研究動物「欺騙」行為的方法似乎只有詮釋擬人論。

最常見的動物欺騙行為大概是偷藏食物。有一種恆河猴在發現食物後，會將食物隱藏在某個位置，不讓其他同伴知道，也不會引導其他同伴靠近牠藏食物的地方。狼一定是在沒有其他同伴在場的時候，才會放置食物或是拿取早先貯藏的食物。黑猩猩甚至會為了要獨吞所有可取得的食物，而給牠的同伴不正確的訊息。

另外更有名的是所謂的折翼偽示。在美國東海岸有一種名叫雌鳩的鳥，會做出一種著實令人讚嘆的行為。雌鳩習慣在海灘或沙丘築巢。像所有動物一樣，牠們的巢都偽裝得很好，以避免被可能的掠食者發現，保護鳥巢裡的幼鳥或者鳥蛋。但是鳥類雖然能將鳥巢偽裝得很好，卻不表示掠食者就不會靠近。雌鳩有一種行為能夠有效地引開掠食者。當母鳥發現掠食者（例如貓）逐漸靠近鳥巢的時候，母鳥會擺動牠的尾巴，然後出現走路很笨拙的樣子，將牠的翅膀拱起來拖著走，或者慢慢飛，就好像翅膀受了傷一樣。牠藉著假裝翅膀受傷，來誘引



掠食者，牠甚至還會假裝一拐一跳的，朝著遠離鳥巢的方向移動。當掠食者被誘引到離開鳥巢夠遠的時候，雌鳩會突然迅速飛高，甩開掠食者。

如果靠近鳥巢的不是掠食者，而是牛或者其他放牧的動物，雌鳩不會出現折翼偽示這種欺敵的行為。不過這不表示牠不會對牛採取行動。牛不吃成鳥、幼鳥和鳥蛋，因此折翼偽示對牛是沒有效的。但是牛卻會不小心踩壞鳥巢。既然「折翼偽示」沒有效，雌鳩發展了不同的作法：在牛逐漸靠近鳥巢的時候，雌鳩會悄悄地靠近牛，然後忽然振翅疾飛，甚至向牛的方向撲過去，讓牛因受到驚嚇而改變行進的方向。

雌鳩還會做出另外一種動作。牠會在某個沙地的位置坐著，緩緩蠕動身體，表現出坐在鳥巢孵蛋的樣子。不過，別被騙了。牠只是裝裝樣子，讓其他生物誤以為牠的鳥巢在那個位置。

你認為雌鳩聰不聰明？你認為這種鳥類有沒有心智？你認為牠有沒有思想？你認為牠有沒有在欺騙？

這些動物真的展現了偽裝和欺騙的行為嗎？即使是，偽裝和欺騙的行為當然是心智的具體表現？不見得！讓我說明我的想法。

首先讓我們回到擬人論的問題。對於這些動物行為的描述當然不是虛構擬人論的作法，因為我們不是在作



文學、藝術的創作；我們並不是在想像卡通和童話裡的動物。對於這些動物行為的描述似乎也不僅僅是詮釋擬人論的作法。大多數人傾向於承認這些動物確實表現出偽裝和欺騙的行為。所以我們現在要問的是：究竟偽裝和欺騙的行為是不是一定出自心智的作用？

如果我內心有意圖要偽裝和欺騙，我的意圖會導致我出現偽裝和欺騙的行為。我的內心意圖解釋了我的偽裝和欺騙。但是反過來說就不一定了。我可以是在無意中出現了偽裝和欺騙的行為。我既無心偽裝和欺騙，我也沒有發現我在偽裝和欺騙。舉例來說，我和幾個朋友一起去杉林溪一帶遊玩。我碰巧穿了一套迷彩裝。當我們在森林中探險的時候，由於我穿的是迷彩裝的緣故，我一下子就「消失」在森林裡。我的朋友看不到我，雖然我正好就站在離他們不遠的前方，抬頭探視樹上的鳥巢。我無意偽裝我自己，我也無意欺騙我的朋友，但我當時確實造成了偽裝的效果，以及朋友誤以為我不在他們前方的錯覺。所以從一個個體會出現偽裝和欺騙的行為，不足以推斷那行為出自於那個個體的內心意圖。同樣地，這些動物雖然出現了偽裝和欺騙的行為，從這點還不足以推斷牠們的行為是出自於牠們內心意圖的——雖然同樣不足以推斷牠們的行為不是出自於牠們內心的意圖。

再者，要具備思想，就一定要具備概念。如果有偽



裝和欺騙的意圖，就一定有「偽裝」和「欺騙」的概念。我們要承認這些動物有這些概念嗎？概念有一個重要性質：它應用的對象相當程度上是不受限制的。舉例來說，「飛」的概念可以用在大多數的鳥類、昆蟲、竹蜻蜓、以及飛機，乃至於想像中的龍。以雉鳩的行為作例子。如果雉鳩具有「欺騙」的概念，雉鳩不應該只針對牠的掠食者和牛馬等產生欺騙的行為，牠應該有可能對於牠的同類、或者牠周遭的其他生物進行欺騙。但是牠的行為只適用於類似的環境，跟牠和牠的下一代的生存有關的那一類環境，卻不適用於其他的環境。所以我們是不是要承認牠具有「欺騙」的概念，還有所保留。

此外，我們應該試試看能不能找到對於動物偽裝和欺騙行為的其他解釋，一種不需假設牠們有心智的解釋。不意外地，生物演化理論似乎能夠提供這方面的解釋。這些動物所表現的偽裝和欺騙行為，是生物為了適應環境而演化出來的。這些動物在牠們生存的環境中，面臨不同的生存困難，而演化出相應的的行為模式來應付環境中的困難。例如，食物來源的稀少迫使狼和黑猩猩必須珍惜得之不易的食物，而珍惜的方式之一是不分享食物。非洲花豹在獵得羚羊之後，會將食物搬到樹上，為的是不讓其他肉食動物（如鬣狗、土狼）爭食（因為鬣狗、土狼不會爬樹）。如果能夠找到適當的生物演化解釋，



或者其他不必用到心智的解釋，我們又何必非要用動物有心智來解釋牠們的行為呢？

「折翼偽示」確實可以承認是一種偽裝和欺騙的行為，但是這種行為可以訴諸生物適應理論來說明，所以還不足以從鴿會展現這類行為，就據以推斷牠們是有心智的。比較有趣的問題反倒是：被「折翼偽示」欺騙的動物會不會出現「被騙了」的行為？甚至會不會出現「反欺騙」的行為？即使對於鴿的研究無法提供我們關於這方面的線索，對於黑猩猩之類、在演化上更接近人類的動物，或許這是值得作進一步瞭解的課題。

我不是主張所有動物行為一定都能訴諸生物演化的解釋，進而否認動物有心智。我不過是在遵循我之前提到的策略：如果能夠使用生物「本能反應」的概念來理解動物行為，就不必使用「制約」概念以及「意向」概念來理解動物行為。我在這裡對於幾個動物行為的討論，只是作為一個引子，希望引起更多的思考。

請問，你覺得動物有心智嗎？  
為什麼呢？





# 結論 人心難測？

——心與認知研究的未來





哲學關於心智的研究我們已經看到了很多，科學要放到哪個位置呢？許多人誤以為哲學跟科學的差別在於哲學是進行概念分析而已，跟科學一點關係也沒有。這是對於 20 世紀哲學，尤其是英美哲學的誤解。概念分析是 50 年代以前的事，主要出現在道德哲學和政治哲學領域等，比較屬於哲學旁枝的地帶。它們的研究主要是對於跟道德和政治有關的概念進行分析。然而它們的發展至今一直不成熟，其末流者更走到枝微末節，陷入文字遊戲的泥淖。不但沒有能力來對現實生活中的道德與政治現實所帶來的哲學問題進行反省，又不瞭解社會科學關於人類種種社會現象以及人性的研究。空談使得這些哲學領域成為毫無貢獻的一門學問。

英美哲學的主流領域則鮮少在對概念進行分析。以形上學為例，哲學中最核心、最根本的議題，都在形上學的範圍裡，它所研究的主題包括等同、本質、共性、自然類、時間……等，但它絕對不是在作概念分析。例如當代形上學對於「時間」的研究，當然無可避免地必須就愛因斯坦相對論來探討。又例如形上學對於「自然類」的研究，當然不是在分析自然類這概念，而是在探討豬、狗、牛、羊……等，之所以自成一類的關鍵。自然類研究的重要性在於，能夠藉以說明科學定律是什麼、解說接受歸納法的道理何在、乃至於進一步提出我們要



如何看待外在的物理世界和生物世界……等問題。

其他主流哲學領域也是一樣。科學哲學的研究，例如科學解釋、理論化約、科學核驗、科學定律、歸納與機率……等，也不是在作概念分析，而是在研究促使我們接受科學的那些重要科學活動以及科學性質。語言哲學追問語詞與句子的真假以及意義的問題，同樣不是在分析概念。



愛因斯坦 (Albert Einstein, 1879-1955) 美國猶太裔科學家。生於德國。一生以文學家追求美的態度和哲學家探索真理的精神，致力於宇宙的思考研究。1905 年發表有關「布朗」運動理論、光電理論、相對論等論文。1921 年獲頒諾貝爾物理獎。

知識論的研究，從傳統知識論到自然主義知識論的發展，引進演化理論以及認知科學的研究成果，來探討人類知識，同樣也不是在分析概念。以心與認知的哲學研究來說，我們到現在為止，還沒有對於哪個概念進行分析過。簡單說，這些主流哲學領域是在對於重要的性質和現象進行哲學思考，以發掘其真象。

尤其重要的，主流哲學的發展跟科學的發展互相影響，使得哲學對於人類的貢獻更為實質，其中尤以心與認知哲學的研究更為顯著。20 世紀對於心智的研究，從純粹哲學的思考，到與計算科學和人工智能的互動、認



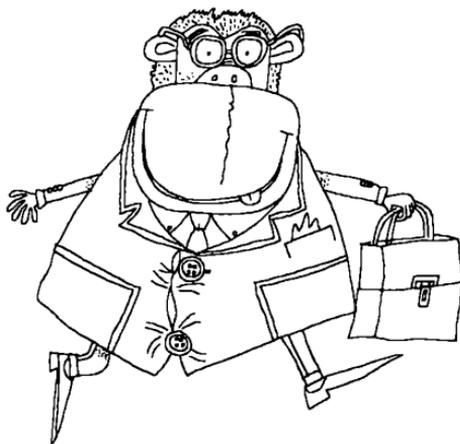
知神經科學的加入、認知科學的興起與昌盛，到認知發展心理學對於幾個星期大的嬰兒，甚至幾天大的嬰兒的認知研究，到語言學對於語言能力的研究、演化心理學引入演化理論來探討心智的問題、認知動物行為學探索動物心智，我們看到的是：哲學家 and 科學家共同思考「心智究竟是什麼」的問題。儘管對於問題的探討與研究方式，哲學和科學的研究明顯有區隔，對於心智研究的重大問題其實已經很難區別究竟是哲學的問題，還是科學的問題。哲學與科學彼此影響頻繁而深入，對於心智的研究見證了哲學與科學如何攜手合作，而獲得了豐碩的成果。我們對於心智的認識也已經比起 50 年前進步非常非常得多。

我們已經粗略看過計算理論、人工智能、行為主義、認知動物行為學等科學研究，在心與認知哲學的研究上所佔有的地位，以及所提出的實質貢獻。未來對於心與認知的研究會往哪個方向呢？這是誰也無法預知的。但是可以肯定一點，哲學與科學對於心智進行二合一的共同研究是一個重要的趨勢。在這方面，目前可以看到的是，認知神經科學以及演化心理學參與了這方面的探討。有愈來愈多的哲學家相信藉由科學的研究成果，我們對於心智的理解將會更深入、成熟。不但如此，科學研究的成果會使得一些哲學問題變成無意義的，而有一些哲



學問題則可能因此需要做一些釐清，甚至於會產生新的哲學問題。這是科學研究回饋給哲學的地方。至於究竟有哪些哲學問題可以獲得解決，有哪些是科學研究的成果解決不了的，則還有待進一步的深思。長遠來看，「意向性」的問題仍然懸而未解，如何藉由對於思想內容的說明來理解「意向性」，仍需哲學的努力。而哲學上另一個艱難的有關意識「主體性」的問題，非常可能是當前的科學無法解決的。若是如此，有些哲學家主張：我們可能需要重新思考「科學」究竟應該具備什麼特徵？將來我們甚至必須發展出新科學，來解決我們目前無法解決的問題。那新科學會是什麼樣子呢？我們發展得出來嗎？

人心難測？抑或我們終有突破的一天？





## 中英名詞索引

- 一元論 monism  
人工智能 artificial intelligence (AI)  
中文房間論證 chinese room argument  
中國頭故事 china head example  
丹尼特 D. C. Dennett  
化約式的 reductive  
化約唯物論 reductive materialism  
巴柏式的心 popperian mind  
心; 心智 mind  
心物二元論 mind-body dualism  
心物平行論 parallelism  
心物因果交互作用原則 principle of mind-body interaction  
心物差異原則 principle of mind-body difference  
心物難題 mind-body problem  
心理內容 mental content  
心理因果原則 principle of mental causation  
心理狀態 mental state; psychological state  
心理表徵 mental representation  
心理現象 mental phenomenon  
心理機制 psychological mechanism  
心腦同一論 the mind-brain type identity thesis  
方法論行為主義 methodological behaviorism  
丘曲冷 Paul M. Churchland



- 主體性 subjectivity
- 他心問題 problem of other mind
- 功能角色 functional role
- 功能論 functionalism
- 可驗證原則 principle of empirical verifiability
- 古典學派 classicism
- 史金納 B. F. Skinner (1904-1990)
- 史賓諾沙 Benedict de Spinoza (1632-1677)
- 布拉克 Ned Block
- 布倫他諾 Franz Brentano (1838-1917)
- 先驗 a priori
- 同構 isomorphism
- 因果角色 causal role
- 因果角色功能論 causal role functionalism
- 多重可具現性 multiple realizability
- 存有論 ontology
- 老百姓心理學 folk psychology
- 自然類 natural kind
- 艾爾 A. J. Ayer (1910-1989)
- 行動測試 action test
- 佛列格 G. Frege (1848-1925)
- 佛列格難題 Frege's problem
- 佛德 Jerry A. Fodor (1935-)
- 完美偽裝者論證 perfect pretender argument
- 折翼偽示 broken-wing display (BWD)
- 私密性 privacy
- 那個樣子 what it is like



- 具現 realization  
卓斯基 Fred Dretske  
取消式的 eliminative  
取消唯物論 eliminative materialism  
命題 proposition  
命題表徵 propositional representation  
命題態度 propositional attitude  
帕南 Hilary Putnam (1925-)  
性質二元論 property dualism  
拉美特利 La Mettrie (1709-1751)  
沼澤人 Swampman  
物理因果封閉原則 principle of physical causal closure  
物理論 physicalism  
知識論 epistemology (theory of knowledge)  
知識論證 knowledge argument  
表象 appearance  
表徵 representation  
表徵假設 representational hypothesis  
表徵論 representationalism  
金恩 J. Kim  
阿姆斯壯 David M. Armstrong (1926-)  
非化約唯物論 non-reductive materialism  
信念—慾望心理學 belief-desire psychology  
思考實驗 thought experiment  
思維語言假設 language of thought hypothesis  
指涉 reference  
指涉隱蔽 referential opacity



## 人心難測

- 洛克 John Locke (1632-1704)  
胡塞爾 Edmund Husserl (1859-1938)  
衍生意向性 derived intentionality  
計算 computation  
計算心理學 / 計算心理學派 computationalism  
計算表徵系統 computational-representational system  
個例 token  
個例物理論 token physicalism  
原初意向性 original intentionality  
弱邏輯行為主義 weak logical behaviorism  
特許管取 privileged access  
納格 T. Nagel (1937-)  
能力假設 ability hypothesis  
馬雷布朗赫 Malebranche  
副現象論 epiphenomenalism  
唯心論 idealism  
唯物論 materialism  
常識心理學 common sense psychology  
強邏輯行為主義 strong logical behaviorism  
理論化約 theoretical reduction  
異例一元論 anomalous monism  
笛卡爾 René Descartes (1596-1650)  
符號操作系統 symbol manipulation system  
透明定理 the transparency thesis  
透明性 transparency  
傑克遜 Frank Jackson  
最佳解釋推論 inference to the best explanation



循環論證 circular argument  
無識人 Zombie  
等同; 同一 identity  
萊布尼茲 Leibniz (1646-1716)  
萊布尼茲定律 leibniz' law  
超級斯巴達人論證 super Spartan argument  
黑白瑪麗故事 the black and white Mary example  
意向性 intentionality  
意向狀態 intentional state  
感質 qualia  
感質缺如 absent qualia  
感質對調 inverted qualia  
概念分析 conceptual analysis  
概念表徵 conceptual representation  
瑟爾 John Searle (1931-)  
葛雷哥式的心 Gregorian mind  
路易斯 D. Lewis  
雷勒 K. Lehrer  
圖林測試 Turing test  
圖林算機 Turing machine  
圖像式表徵 iconic representation  
實在; 實在界 reality  
實體 substance  
實體二元論 substance dualism  
管取 access  
算機功能論 computational functionalism  
維根斯坦 Ludwig Wittgenstein (1889-1951)



語言式表徵 linguistic representation  
語言測試 language test  
語意組構性 semantic compositionality  
語意學 semantics  
認知 cognition  
認知主體 cognitive subject  
認知科學 cognitive sciences  
動物認知行為學 cognitive ethology  
認知構造 cognitive architecture  
德威笙 D. Davidson (1917-2003)  
範疇謬誤 category mistake  
論證 argument  
質性的 qualitative  
機遇論 occasionalism  
隨附 supervenience  
擬人論 anthropomorphism  
擬似意向性 as-if intentionality  
雙生地球案例 the twin earth example  
雙面向論 dual aspect theory of mind  
羅素 Bertrand Russell (1872-1970)  
類比推論 argument from analogy  
類型 type  
類型物理論 type physicalism  
邏輯行為主義 logical behaviorism  
邏輯實證論 logical positivism  
驗證 empirical verification  
觀察語句 observational sentence