「強AI」論旨及其問題的性質

● 馮耀明

《二十一世紀》於1992年6月號刊出了李逆熵先生的大文〈我看人工智能辯論〉，李先生深入淺出地介紹了最近西方對人工智能的爭論，特別提到西爾 (John R. Searle) 和彭羅斯 (Roger Penrose) 二人對「強AI」的駁論。我是從事中國哲學研究的學者，但對西方心靈哲學的新發展（如認知科學(cognitive science)）頗感興趣。讀過李先生的文章後，覺得有所啟發。但對若干問題的理解與李先生頗不相同，故特撰一短文就教於李先生。

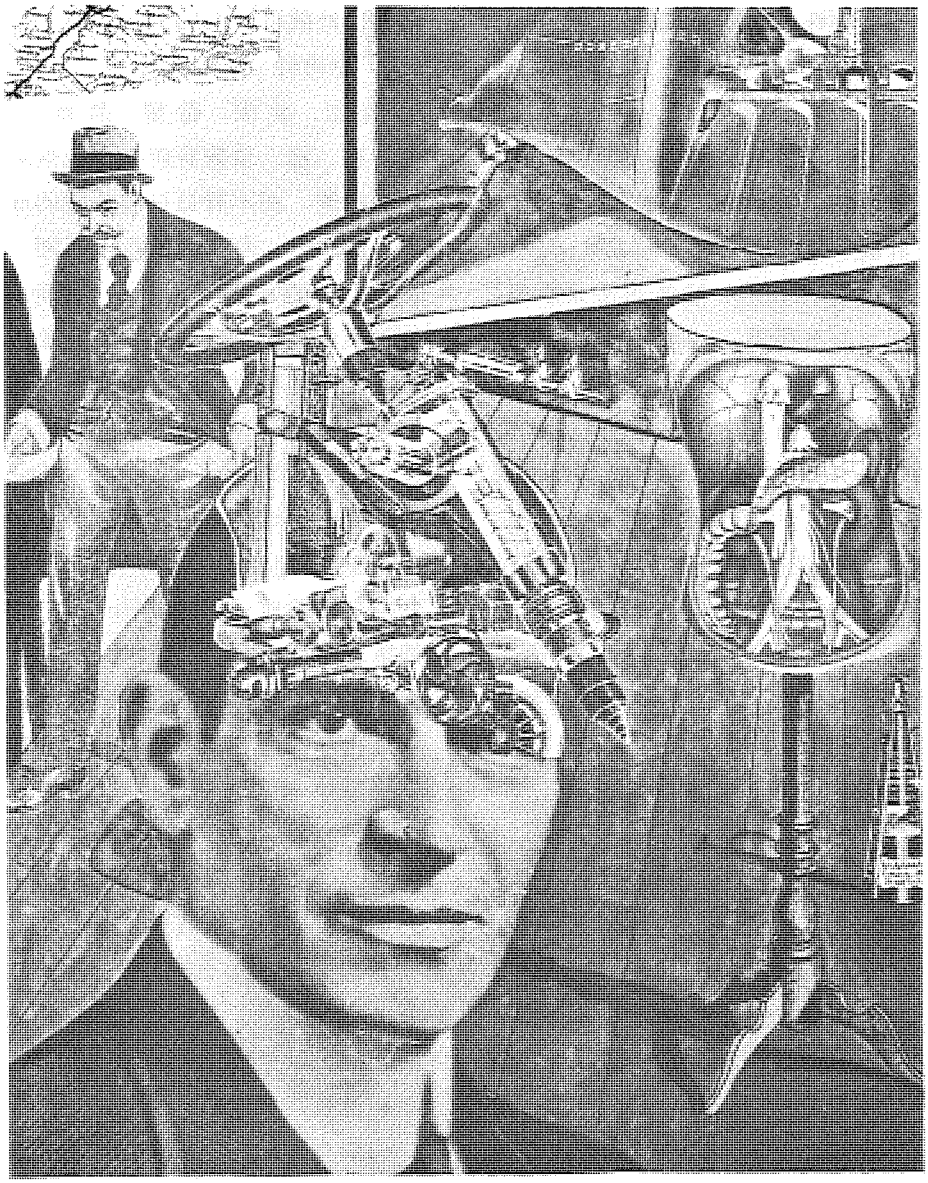
強AI是甚麼？

第一個問題是有關「強AI」論旨的問題。依照李先生的說法，「有關AI的重大爭論，針對的是『強AI命題』。這一命題認為，人類終有一天能够造出一副在認知和思維能力方面皆與人完全無異（如果不是更優勝）的機器。」（李文，頁60）。

我們知道，「強AI」這一概念是由西爾首先提出來的。但西爾對「強AI」的說法，與李先生的說法有極大的出入。依照西爾的說法①，「Strong AI」是「一個藉着設計程式來重造及說明心靈事物之計劃」②：其論旨是「心靈過程是在形式地界定的成分之上的計算過程」③：較明顯的說法是：「具備恰當程式的電腦實質上就是心靈，意即電腦給與正確的程式，我們就能說它實際上會理解以及有其他認知狀態。」④不過，我認為最能突顯「強AI」論旨的，是這篇經典性文章和西爾其他相關著作所一再強調的口號：「心靈之於大腦，一若程式之於電腦硬件。」類似的說法有：

> 大腦僅只是數碼的電腦，而心靈僅只是電腦的程式⑤。
>
> 我把此一認爲具有一心靈之所有即是具有一程式的觀點，叫做強AI⑥。

由此可知，西爾反對「強AI」，是反對以「心靈等同於程式」（以「心靈



「意向性」不僅指涉意識，也指涉各種無意識的心靈狀態。

狀態或過程等同於計算狀態或過程」）的觀點，而不是如李先生所說的，反對「人類終有一天能够造出一副在認知和思維能力方面皆與人完全無異的機器」。此種誤解在西方學術界也不是沒有出現過的，例如Philip Cam認為「強AI」是「人工地創造意向性的任何試圖」⑦，而省略了「只藉設計程式」一語，乃是對西爾「強AI」論旨的極大誤解。即使在 *Scientific American* ⑧上與西爾對辯的 Paul M. Churchland 和 Patricia Smith Churchland夫婦，也誤解了他們之間的論爭所在。西爾指出：他們二人以為「強AI」是主張電腦終有一天能產生思考，而西爾在常識的基礎上否定此可能性。但西爾認為這既不是「強AI」的主張，而他的論證也與常識無關⑨。因為，反對「強AI」並不表示否定電腦或其他機器有一天可能會產生思考，而只是基於邏輯及概念分析的結果，否定電腦或其他機器只靠體現語法程式便能產生思考，及否定電腦的程式等同於大腦的心靈之說法。西爾認為：如果「機器」被了解為能達至某些功能的物理系統，人類便

是一種特殊生物種屬的機器，由於人類能思考，所以有些機器也能思考⑩。如果問題是：「一個人造物能思考嗎？」答案要視乎這是一個怎樣的人造物。西爾認為：假如我們能够設計出一部機器，與人類在分子與分子之間毫無分別，如能再造其原因，便能再造其後果。因此，這答案無疑是肯定的⑪。但如果問題是：「一個人造物體現恰當的電腦程式，有恰當的輸入和輸出功能，是否足以產生或構成思考？」則答案是否定的。因為，思考不僅是操作一堆堆無意義的符號，它包含着具有意義的語意內容。電腦程式既被界定為純粹形式的語法，那麼這種電腦人造物依定義是不能思考的⑫。

關於意向性的論爭

思考不僅是操作一堆堆無意義的符號，它包含着具有意義的語意內容。電腦程式既被界定為純粹形式的語法，那麼這種電腦人造物依定義是不能思考的。

西爾反駁「強AI」的論證，牽涉到語法與語意差距的問題，以及與語意內容有關的意向性的問題。現在我們要與李先生討論的第二個問題，便是「意向性」一概念的理解問題。李先生認為「意向性」一詞是「整場AI爭論的核心」，這無疑是十分確當的論斷。但他認為對應於「意向性」一詞，「我們常用的字眼，則是『自我意識』(self-consciousness)和『自由意志』(free will)。」(李文，頁59)此一理解並不準確。雖然「意識」與「意向性」及「自由意志」都有相關之處，但三者並不一樣。「意向性」既不能包含「意識」和「自由意志」二者，更與「自我意識」無關。

依照西爾的規定，「意向性」是指心靈狀態的一種特性，藉着它，我們的心靈狀態指向、涉及或指涉此心靈狀態以外的世界上的對象和事態。依此，「意向性」不僅指涉意向，它也指涉信念、欲望、希望、恐懼、愛、恨、慾、惡、羞等有所指涉或指向的心靈狀態，而不管這是有意識的或是無意識的⑬。意向性與意識不同，因為有許多意識的狀態不是有意向的，例如突如其來的驚喜；而許多意向的狀態也不是有意識的，例如我有許多信念是目前沒有想起的，而且可能永遠不會想到的⑭。

有些人認為意識狀態也就是一種自我意識的狀態，西爾則認為這種說法不是多餘地真(trivially true)，便是簡單地假。依照「自我意識」一詞的一種特別的用法，由於我們通常能够把注意力從意識經驗的對象，轉移到此一經驗本身，這可以說「任何意識都是自我意識」。但照一般日常的用法，「自我意識」是指對某一意識狀態本身之意識，則並非所有意識狀態都是自我意識的⑮。退一步來說，即使意識就是自我意識，由於有的意向性是無意識的，可見「意向性」與「自我意識」在概念上並不相同。

李先生說：「西爾強迫『強AI命題』的擁護者面對這樣一個問題：姑勿論某副機器能否真的通過『圖林試驗』，但你們認為機器真的能够擁有思想、感情，亦即擁有自我意識嗎？」(李文，頁61)一如上述西爾的說法，有些思想信念和感情狀態不是有意識的，即使是有意識的，也不一定是有自我意識的，因此這裏的「亦即」二字用得並不恰當。無論如何，李先生認為「大部分的『強AI』擁護者對上述問題的答案其實是肯定的」，「只是他們一向不願意宣揚這一觀點。……因為『自我意識』這回事按其本質是永遠無法確立的。」對電腦來說，這

極可能只是「一些極其巧妙的程序所產生的結果」。(李文，頁61)此一說法正是大多數「強AI」論者的化約論的觀點，以為可以把有意向或有意識的心靈狀態化約為非心靈的狀態(例如計算狀態)，把高層次的語意現象透過一種遞歸的分解程序歸約到最低的語法層次。西爾認為這種說法是行不通的，他認為如果沒有一個縮形人站在遞歸分解的程序之外，我們甚至沒有一套語法可供運作。因此，這種"homunculus fallacy"是消除不了的⑯。其次，西爾最近出版的新書*The Rediscovery of the Mind*更指出：把意向性或意識狀態化約為其他東西的主張，乃是一種以「第三身」取代「第一身」的觀點。由於「第一身」的經驗是真實的，把「第一身」的經驗內容轉化為「第三身」的經驗內容，自然會是"something else"(chs. 1–2)。因此，Jerry Fodor、D.C. Dennett及W.G. Lycan等人的「意識」概念並不等同於西爾的「意識」概念，「意向性」一概念亦如是(同上，頁51、55)。(李先生說：「我固然知道我自己存在，亦即我有自我意識。」(李文，頁61)這似乎是「第一身」的觀點。但上文說「『自我意識』這回事按其本質是永遠無法確立的」，以及下文「一些極其巧妙的程序所產生的結果」之說法，則有從「第三身」的觀點看「自我意識」之傾向。而前者似「消除主義」(eliminativism)的立場，後者似「化約論」(reductionism)的立場。)

至於「自由意志」一概念，乃牽涉心靈自由地選取行為，抑或行為完全為外界所因果地決定或命定的問題，此一概念的內涵明顯地超出了「意向性」和「意識」的意指，因此也是需要加以分疏，不可混為一談的。

兩個層次：邏輯與事實

我們要與李先生討論的第三個問題，是「強AI」論旨之問題的性質。「『強AI』論旨是否成立？」這個問題屬於甚麼性質的問題呢？李先生為了說明「意向性」一詞為「整場AI爭論的核心」，他引用了西爾在"Minds, Brains, and Programs"一文最後一段話：「無論大腦如何產生意向性，這種過程必不等同於一項電腦程序。因為單靠電腦程序本身，絕不足以產生意向性。」可是，李先生跟着又說：「歸根究柢，有關AI的最大爭論是：我們終有一天可以造出一副擁有自我

把意向性或意識狀態化約為其他東西的主張，乃是一種以「第三身」取代「第一身」的觀點。

每一事物都是數碼電腦，大腦如是，機器也如是？

意識和自由意志的機器嗎？」(李文，頁59)這似乎是一種滑轉，把原來「單靠電腦程序本身是否足以產生意向性」的論題，轉變為「能否造出一副擁有自我意識和自由意志的機器」的論題。此一滑轉，無疑已使問題的性質改變了。李先生的文章跟着區分「強AI」和「弱AI」命題，他的「強AI」說法與西爾的大異其趣，也許正是由於這一滑轉。

這一滑轉使原先的問題變了質，由一個邏輯及概念性的問題，轉變而為一個科學事實的問題。西爾認為：「強AI」論旨是否成立之問題，不必等待電腦科技之改進。他之反對「強AI」，是完全獨立於任何科技狀態及條件的，這論爭只與數碼電腦的定義本身有關⑰。因為程式只是純粹形式或語法地界定的，而內在心靈狀態依定義是有語意內容的，故程式與心靈並不等同，這是由邏輯分析所致，而不是由科學發現所得的結論⑱。換言之，試圖只藉設計程式來創造心靈的計劃，從一開始便注定失敗。因為這只是一邏輯論證的結果，與科技的發展或程式的複雜性(如由線性發展為平行網絡模型)等技術條件無關⑲。順着西爾的想法，我們可以說：電腦的硬件有特殊的物理因果力量，但程式作為純粹形式的語法，它是沒有任何物理特性的。因此，要問某種硬件配合上軟件(程式或算法)能否產生思考，這可以是一個科學的問題，因為這牽涉到事物的物理性質。但要問軟件本身能否產生思考，這便是一個邏輯及語意上的問題，與科學及技術並不相干。由於「強AI」論旨是後一問題，我們沒有理由把它當作科學及技術問題來處理。

某種硬件配合上軟件能否產生思考，這可以是一個科學的問題，因為這牽涉到事物的物理性質。但軟件本身能否產生思考，是一個邏輯及語意上的問題，與科學及技術並不相干。

西爾認為：不只「電腦只靠其程式便能產生思考」或「心靈就是電腦的程式」(「強AI」論旨)不是科學探究的事實問題，而且「大腦就是數碼電腦」(「認知主義」(cognitivism))也不是科學探究的事實問題⑳。在1980年的經典性文章中，西爾認為：「大腦當然是數碼電腦，因為每一事物都是一數碼電腦，大腦也是。」㉑彭羅斯認為事實並非如此㉒。但彭羅斯似乎忽略了西爾的原意。西爾說「每一事物都是一數碼電腦」，是基於「任何事物都可以被描述為(described as)一電腦程式的體現」之提法㉓。更清楚的說法是：「從數學的觀點看，不管任何事物都可以被描述為好像(described as if)是一部數碼電腦。」即使桌面上的筆也可被如此解釋㉔。這種說法其實是來自Ned Block的觀點，他認為依計算的說法，我們能從貓、鼠和芝士、或槓桿、或水管、或白鴿、或任何其他東西，製造出與你和我相似的「大腦」，只要兩個系統是「計算地等值的」(computationally equivalent)㉕。西爾同意這個觀點，因為他相信形式語法或計算的性質不是物理事物所本有而內在的性質，而是行為者或觀察者透過其解釋(interpretation)而賦予的性質㉖。因此，即使彭羅斯如何努力地在邏輯分析之外，用科學的方法來發掘大腦的實際運思程序並非電腦的算法程序這一「事實」，但也不能否定任何事物可被解釋成電腦這一說法。

西爾反對「強AI」，理由是語意並非內在於語法；他反對「認知主義」，因為語法不是內在於物理㉗。「大腦就是數碼電腦」是一個含糊的主張，西爾從前認為這是多餘的真句，現在則認為它連假都說不上㉘。無論如何，西爾認為「強AI」和「認知主

義」的論題都不是科學事實的問題。不同情「認知主義」的彭羅斯和Hubert L. Dreyfus好像都把這個問題當作是直接的事實問題，但西爾認為他們似乎並沒有對問題的性質有所疑惑㉙。

如果「強AI」及「認知主義」的論題是一些有待科學發掘的問題，則反對「強AI」及「認知主義」的論者（包括彭羅斯本人），便不可以用「皇帝的新心靈」來形容電腦中的所謂「心靈」。因為，一個有待科學發掘的問題，並不是一個子虛烏有的問題。只有像西爾一樣，把這些問題當作是邏輯及概念分析的對象，才可以把電腦中的所謂「心靈」分析出來。

對於李先生所信奉的「人類終有一天可製造出一副擁有自我意識的機器」之理想（就其邏輯的可能性而言，相信沒有多少人會否定此一理想），我雖然沒有資格表示樂觀或悲觀，倒也願意「樂觀」其成。但作為一個哲學工作者，總不免有些哲學上的疑慮。我的疑慮是：如果沒有在概念上釐清甚麼是「意識」，我們又如何能夠將「意識」的研究放到一個堅實的科學基礎之上，又如何能夠進一步去創造「機器意識」呢？

註釋

① Roger Penrose在*The Emperor's New Mind* (Penguin, 1991)中，對「強AI」的用法也是依從西爾的說法。可參閱頁17–23；頁407；頁429；頁447。

②③④㉑㉓ "Minds, Brains, and Programs", *Behavioral and Brain Sciences*, vol. 3, no. 3 (1980), p. 423; p. 422; p. 417; p. 372; p. 268.

⑤⑰⑱⑲㉔ *Minds, Brains and Science* (Penguin, 1989), p. 28; p. 30; p. 31; p. 39; p. 36.

⑥ "Is the Brain a Digital Computer? —Presidential Addresses", *Proceedings and Addresses of the American Philosophical Association*, vol. 64, no. 3 (Nov., 1990), p. 22.

⑦ "Searle on Strong AI", *Australasian Journal of Philosophy*, vol. 68, no. 1 (March, 1990), p. 103, p. 106.

⑧ *Scientific American*, vol. 262, no. 1 (Jan., 1990), pp. 25–37.

⑨⑩ "Is the Brain's Mind a Computer Program?", *Scientific American*, vol. 262, no. 1, p. 28; p. 26.

⑪⑫⑬ *Minds, Brains, and Science*, p. 35; p. 36; p. 16.

⑭ *Intentionality* (Cambridge, 1983), p. 2.

⑮ *The Rediscovery of the Mind* (MIT, 1992), pp. 142–43.

⑯⑳㉕㉖㉗㉘㉙ "Is the Brain a Digital Computer?", pp. 28–29; pp. 21–22; p. 25; pp. 26–27; p. 27; p. 35; p. 24.

㉒ *The Emperor's New Mind*, p. 23.

馮耀明 現為香港中文大學哲學系講師，此前曾任新加坡東亞哲學研究所研究員。主要著作包括《卡納普與邏輯經驗論》、《中國哲學的方法論問題》，編有《分析哲學與科學哲學論文集》（與劉述先教授合編）、《無慚尺布裹頭歸——徐復觀最後日記》（與翟志成合編）。