

第三章 博之論自我知識

一般來說，我們通常不需要透過觀察或推論的方式，就能知道自己在想什麼。但是當我們要知道其他人在想些什麼的時候，我們往往需要透過一些證據或者根據證據的推論——比方說他人的話語或行為——才能得知。我們的語言行為表現了這種差異。一個人 A 以第一人稱的方式對自己做出的心理歸屬，比方說「我認為(think)今天的天氣不錯」，通常會被 A 的聽眾理所當然的認為是真的。這個意思是說，在沒有其他理由懷疑 A 在說謊或者是非理性的情形中，A 的聽眾一般認為 A 的自我歸屬反映了他的心理狀態，而且 A 知道他的話語在表達什麼。相對的，當其他人說他們認為 A 在想什麼的時候，聽眾通常不做這樣的推想。因為，其他人要透過理解 A 的話語或行為來瞭解說話者的心理狀態，而他們或許會誤解或一知半解 A 自我歸屬的內容或者他的行為表現，因而對 A 做出錯誤的心理歸屬。但我們通常推想說，我們沒理由說 A 不知道自己在講什麼、想什麼。

這些現象表明了，一般在理解一個說話者第一人稱式的自我歸屬時，我們會認定說話者具有一種權威性的地位，一定知道自己在表達什麼，不會產生像聽眾可能會有的那種誤解；在這個意義下，說話者比起他的聽眾來說，更知道自己的話語的意義。總結來說就是：一思想者對自己的心理狀態具有直接的、非推論性的自我知識(self-knowledge)，這種自我知識具有其他人稱式的心理歸屬所沒有的權威性格。這些特徵讓自我知識又被稱為權威的自我知識(authoritative self-knowledge)或者第一人稱權威(first person authority)。

權威性自我知識與他心知識之歸屬之間的這種不對稱性，需要哲學上的解釋。不過在提出解釋之前，我們的焦點暫且轉向心理內容的外在論。心理內容的外在論主張，一思想者之心理狀態的內容，部分是由思想者之外的外在事物——例如物理環境或語言社群——所決定的。而乍看之下，外在論很容易被理解為蘊含這樣的主張：要知道我們的心理內容為何，我們必須先知道相干的外在事物為何。在這樣的理解下，外在論與自我知識的觀念似乎有著本質性的衝突：自我知識指的是我們對自己的心理狀態與內容具有直接的、非推論的知識，但外在論主張，要知道自己的心理內容，思想者要先透過經驗與觀察的方式得知相干的外在因素

爲何；若外在論爲真，那麼我們不可能對自己的心理內容（與心理狀態）具有非推論的自我知識；相對的，若上述的自我知識觀念爲成立，那麼我們就不能接受外在論對心理內容的說明。以下我們簡稱這樣的觀點爲「不相容論」(incompatibilism)。

作爲一個內容的外在論者，博之指出，上述自我知識與外在論的衝突只是表面上的。這個表面衝突來自於對外在論的誤解；只要澄清這些誤解，就可以看出外在論與自我知識的觀念是相容的。戴維森(Donald Davidson)以另一種方式指出了同樣的結論，不過他卻認爲博之的外在論本身——尤其是社會外在論——會有損我們具有權威性自我知識的觀念¹⁷。本章我們將以這兩個問題，檢視博之的外在論與自我知識觀念是否有本質上的衝突。第一節，我們先討論戴維森所提出的批評，藉此看出博之容易引起不相容問題的地方。第二節，我們將引述博之對他的思想實驗所進行的辯護。第三節，我們將試著爲博之提出辯護，並顯示出他的立場沒有戴維森所批評的問題。第四節，我們要討論博之對自我知識的看法，以及他對於表面衝突所做出的診斷。透過這些討論，我們要初步的解消掉外在論與自我知識觀念的表面衝突。

第一節 戴維森對社會外在論的批評

博之的外在論主張旨在反對各種心靈哲學中的個體論假設，尤其是反對笛卡兒式的個體論。他對於笛卡兒式的心理模型描繪如下：主體對心理內容的內省，被當作一種直接的、立即的經驗；而主體在思考什麼，是由主體理解力的力量與範圍，以及他對於內容的內在傾向來決定。這種模型蘊含說一個人可以透過內省達到對於心理內容的完全理解，而只有他完全理解的內容才能被歸屬給他(IM, pp.60-61)。不過，我們之前討論過的關節炎思想實驗，直接指出相反的現象：一說話者（病人 p）對關節炎有所誤解，而我們仍能夠以社群（專家或其他人）的標準，也就是以字面意義(literal meaning)的方式來詮釋他的話語，歸屬給他相應

¹⁷ Davidson, D.(1987), 'Knowing Our Own Mind', in *Subjective, Intersubjective, Objective*. U.S.: Oxford University Press, 2001, p.26. 以下簡稱本文爲'KOM'。

的心理內容。根據前面的分析，我們習慣以社群的標準來詮釋病人 p 的自我歸屬，並根據（作為標準的）專家對於關節炎的認識，來測定 p 對於關節炎的認識程度。可以明顯的看出來，病人 p 無法僅僅透過對自己思想的仔細反省，來知道說「關節炎」一詞的正確用法。但這並不妨礙我們透過他表現出一定的語言能力與隱然導向社群標準的調整，確定出他具有關於關節炎的想法，並確定出其中一些想法是錯誤的。博之評論說：

…[思想實驗]顯示出，一個人的思想內容並非由他身上發生的事情，或者由他透過仔細反省就可以得到(accessible)的事情來確定。…內容有時對內省來說是「不可得的」，這單純只是因為許多心理歸屬並不預設主體對他的思想內容具有完全的掌握(fully mastered)。 (IM, pp.62-63)

博之的外在論雖然有道理，但乍看之下，這種說法似乎是暗示說我們不一定總是完全清楚自己在想什麼。戴維森(1987)針對博之對於笛卡兒模型的批評，提出了這樣的質疑：似乎按照博之的主張，當我們搞錯、誤會或者一知半解某個語詞的公共意義(public meaning)的時候，我們就會搞錯、誤會或一知半解自己使用這個語詞所表達出來的任何信念內容(KOM, p.26)。這明顯與我們認為我們一般都具有權威性的自我知識的看法相衝突，因為我們在說話的時候，通常相當清楚自己在講些什麼、想些什麼。即使我們使用的語詞被指出是誤用的或者不恰當的，通常我們也很清楚自己想要用這些語詞指什麼東西、表達什麼樣的想法。

戴維森雖然也支持博之的反個體論立場，並且相當同情博之的物理外在論¹⁸，但是他認為社會外在論似乎有損我們的權威性自我知識，並認為有必要質疑博之的論證前提。戴維森並不接受語言意義（以及相應的概念）是由社會的用字標準來決定的；他認為社群的標準用法並沒有如博之所認為的那麼重要。他舉出一個例子來說明這點：假設 S 與 R 兩人都相信 Y 有關節炎（Y 也真的有關節炎），而 S 以為關節炎只包括由鈣質堆積在關節處造成的病症。設想 S 與 R 都向 Q 報導他們的信念：「我認為 Y 有關節炎」。我們可以想見，如果 Q 知道 S 與 R 的相干想法，卻只向 Y 報導說：「S 與 R 都相信你有關節炎」，一定會誤導 Y 對於 S 的思想歸屬。戴維森認為一個正確的作法是，Q 有必要加一句補充說「S 與 R 都

¹⁸ Davidson, D. (1990), 'Epistemology Externalized', in *Subjective, Intersubjective, Objective*. U.S.: Oxford University Press, 2001, p.200. 以下簡稱'EE'。

相信你有關節炎，只是 S 認為關節炎一定是由鈣質堆積引起的」。這個補充顯示出，Q 對於 S 與 R 的思想歸屬其實是不同的，也可以顯示出兩人用同樣的一句話（「我認為 Y 有關節炎」），其實表達了不同的信念內容(KOM, p.27)。

這個例子顯示出，如果我們僅僅按照字面意義的方式來理解 S 與 R 的話語，那麼我們肯定會誤解他們具有同樣的信念。戴維森的補充則顯示了，有時候我們以說話者意義來理解說話者的話語，才是正確的作法。我們不一定總是要順從社群的標準，才能恰當的理解說話者的話語。

戴維森的想法是：

…日常的意義和態度歸屬奠基於大量的關於什麼存在、什麼不存在的模糊假設上面，這些假設由歸屬者、被歸屬的人以及歸屬者的聽眾所分享。如果有些假設被證明是錯的，那麼我們就修改我們用來報導的字詞。通常我們以很懶的方式這樣做：我們將他人的話信以為真，即便這些話並未完全反映出說話者的思想或意義的一些面向。…當然博之不一定會對此讓步，因為他可以堅持此報導是字面上正確，但就像任何的信念一樣，都可能被誤導。另一方面來說，我認為這樣的回應忽略了一個信念的內容必然奠基於其他信念內容的地方；思想並非獨立的原子，所以不可能有任何簡單、嚴格的規則可以用來作單一的思想的正確歸屬。(KOM, pp.27-28)

從上面的例子以及引文可以看出，戴維森強調的是詮釋活動的另一個面向：一思想者的思想內容，部分是由思想者其他的相干信念來決定的，如果其中的某些信念被證明是假的，思想者就會修改報導信念的語詞的用法。以上面的例子來說，我們可以推論出，S 在未被指出錯誤之前，他使用「關節炎」指涉的是只由鈣質堆積在關節處引發的病症，而非一般醫生或 R 所謂的關節炎；在相干的信念被證明為假之後，S 會修改「關節炎」用法，指稱一般醫生或 R 所謂的關節炎。

問題在於，有沒有可能說 S 有這些的信念沒錯，但 S 對於自己的想法沒有完全的理解，以致於 S 有可能誤用自己的語詞而不自知？對此，戴維森提出自己的詮釋理論，指出這是不可能的。不過，回答這個問題之前，我們要先約略說明戴維森如何解釋我們的語詞意義與心理內容的決定因素。戴維森認為語詞意義（或其表達的心理內容）要以兩個說話者彼此的互動以及兩人與世界的互動，三方面共同決定。他舉例說，假設有一個說話者以及一個想瞭解說話者在說什麼的詮釋

者。再設想每當老鼠在視線良好的情況中出現時，說話者就用手指著老鼠發出：「Ratón」的聲音；而如果老鼠出現在視線不佳的情況中、或者說話者心不在焉的時候，說話者對老鼠的反應就沒有那麼肯定（可能是發出「Ratón？」）。戴維森認為，詮釋者一般都會將說話者所說的「Ratón」理解為有老鼠出現了，除非有很多證據能反駁這樣的詮釋。一個比較合理的詮釋策略是：每次老鼠出現，就導致說話者有意地發出同樣的「Ratón」表述，並且是以肯定的語氣來這樣表達。對兩個人來說，「Ratón」就意指老鼠。當然，詮釋者仍有誤解說話者的可能，不過在詮釋者沒有出錯的情況中，詮釋者的詮釋都是正確的(EE, p.196)。

在這樣的決定模型中，詮釋者是否能懷疑說話者會不會搞錯自己的話語在意指什麼呢？戴維森認為不可能。其理由在於，在詮釋活動的過程中，「除非預設說話者知道他自己意指為何，及知道如何正確使用他所具有的語言，否則就不會有東西可以讓詮釋者加以詮釋」(KOM, p.38)。他指出語詞與概念的正確性標準部分由思想者的相干信念決定之外，部分還要取決於說話者與詮釋者的互動關係：「一個說話者如果希望能被理解，他就必須意圖讓自己的話語以某種方式來被詮釋，如此就必定要意圖去提供他的聽眾一個線索、使他們達成意想中的詮釋。可學習性(learnability)與可詮釋性(interpretability)的必要，才是不可化約的社會因素、才顯示出一個人為何由於他的話語無法正確地被他人解釋，而無法用他的話語來意指某物。」(KOM, p.28)

詮釋活動，預設了說話者知道自己的話語意指為何。在詮釋活動中，詮釋者由於必須透過他的對說話者的觀察與推論，才能理解說話者的話語的意義，這裡就存在了出錯的空間。但是詮釋者必須預設說話者不會搞錯他的話語在意指什麼，否則他無法透過這些線索來理解說話者在表達什麼。不過戴維森指出，詮釋的活動若是失敗，亦即說話者一直無法透過自己的行為表現，讓詮釋者將他的話語與他所要意指的事件連接起來，那麼我們或可認定說話者不知道自己在想什麼¹⁹。

從上述的討論來說，戴維森顯然認為詮釋活動中的確存在有所謂的「社會因素」，不過這種社會因素並非博之所設想的社群常規，而是兩個人在溝通活動之

¹⁹ Davidson, D. (1984), 'First Person Authority', in *Subjective, Intersubjective, Objective*. U.S.: Oxford University Press, 2001, p.13

中產生的一種習慣與默契，使得兩人可以彼此學習與詮釋對方的語言意義。他認為博之的說法會導致對權威性自我知識觀念的威脅，是因為按照博之的說法，說話者使用的話語或概念的正確性標準，總是由其所屬的社群標準來決定，那麼一說話者就常常會被說成是搞錯、誤解或一知半解其心理內容。然而根據戴維森對第一人稱權威的解釋，詮釋活動預設一說話者一定知道自已的話語在表達什麼。根據他的詮釋模型，一說話者的工作就是以詮釋者能理解的方式，使得自已的話語能夠被詮釋；說話者不一定總是要遵循著某種社會規範來講話，才能使詮釋者理解自已的話語。在這個意義下，我們就不會說一個說話者常常搞錯或一知半解他自己在『想些什麼』。

而除此之外，他還明確地指出了，讓他不同情社會外在論的兩個原因：第一個原因是，他認為博之指出我們一般偏好以其他人的說話習慣來詮釋說話者的話語的直覺，是虛假的。理由在於，一方面，要決定出社群的標準為何，是有問題的；另一方面，戴維森認為「當我們以說話者意圖被詮釋的方式來詮釋他時，我們才是最適當地理解一說話者；比起說我們假設他意指或思考的是，另一個”正確地”使用同樣語詞的人所意指或思想的事情，這將會更恰當地解釋他的行為」。(EE, p.199)

第二個原因則是，戴維森並不信任博之『假想在事實上從沒有發生過的情況下、說話者會說些什麼』的思想實驗；他認為他的外在論觀點是根據我們平常在實際的溝通活動中會發生的事情(Ibid.)。

第二節 博之對幾種重新詮釋策略的反駁

以上一節的討論可以看出，戴維森似乎認為如果我們說一個說話者有什麼樣的想法，但是又說他對其中某個觀念是一知半解，那麼就等於說他不完全清楚自己在想些什麼。博之的社會外在論會導致這樣的結果，其關鍵就在於博之認為在詮釋活動中，我們偏好以字面意義進行詮釋與心理歸屬，並以此評估說話者的用字方式與概念應用是否正確。然而，戴維森所改寫的關節炎例子則顯示出，我們不一定總是要以字面意義來詮釋說話者的話語，然後將說話者的關節炎概念說成是一

知半解或有所誤解的；戴文森主張，以說話者意義來重新詮釋(reinterpret)一遍說話者的話語，更能恰當的理解說話者的意圖，也可以讓我們推想說話者完全理解自己在說什麼。然而這樣的批評是否合理呢？

實際上，博之已經預料過，反對者或許會對證成社會外在論的思想實驗提出這樣的反駁：在思想實驗的第一步中，病人 P 所用的「關節炎」一語可以透過其他方式來詮釋，讓我們發現其實他的用法和醫生不同，如此，P 的信念就不能被說成一種誤解。博之試著提出以下幾種重新詮釋的可能主張，並一一加以駁斥，以顯示出他主張的詮釋策略是合理的。(IM, pp.46-55)

第一種方式，是將 P 的心理歸屬詮釋為關於某種東西的 *de re* 態度，也就是說完全以「關節炎」一詞在當下的脈絡中指涉什麼，來詮釋 P 的態度。比方說，將 P 相信他大腿上有關節炎的信念，重新詮釋為他相信某些非關節炎的類風濕疾病在他的大腿上。但博之認為，這種說法忽略了原來的信念是 *de dicto* 的，也就是說「關節炎」一詞出現的方式是在描繪 P 的主觀觀點：他認為關節炎以某種方式發生在他的大腿上。用 *de re* 態度來重新詮釋 P 的態度，會忽略掉 P 的思想的主觀面向，這會把實際情況中的 P 與相反情況中的 P' 之間的差異給取消掉。

第二種方式，是笛卡兒式的回應。在不完全理解的情況中，主體的態度或內容就是不確定的。但博之認為這並不適用於他的思想實驗，因為顯然 P 與醫生 D 都準確的知道與同意，如何去印證或修正(confirm or infirm)P 的信念。如果 P 與醫生都不清楚 P 在想什麼，亦即不瞭解 P 所說的話的真值條件，他們將無法去印證或修正 P 的信念。

第三種方式是歸屬 P 一個對象層次(object level)的信念，也就是歸屬一個只掌握到誤解的觀念，將 P 的信念解釋成正確的。這是說在 P 的誤解中，P 認為「關節炎」一詞可用在關節炎的病症、也可應用在他大腿上的類風濕症上；我們或可用大腿炎這個概念來涵蓋這樣的想法，解釋說其實 P 在提到「關節炎」時其實是在表達大腿炎概念。然而博之認為這種說法有幾個問題：(1).一般我們發現一個人曾經誤解一個語詞時，我們並不會重新詮釋包含這個語詞的心理歸屬（比方說醫生不會用大腿炎概念來詮釋 P 具有的態度）。而且我們常常把此人看作是與專家分享著同樣的態度。(2).假設就算有「大腿炎」這樣的詞彙涵蓋病人的大腿炎概念，這些詞彙與觀念的應用條件其實並不清楚。實際上，沒有任何可識別的判

準可以顯示出這個詞彙正確性標準²⁰。(3).歸屬「大腿炎」一詞的作法是一種特設假說(ad hoc)，就像「燃素」這種詞彙所表示的觀念一樣，為了解釋現象而造出新字、新概念。我們並不會因為發明了一個新字就比較理解 P 的想法，比較簡單且有啟發性的作法，就是去解釋 P 認為關節炎可以發生在大腿上。(4).最重要的是，歸屬對象層次的信念無法說明病人 P 相干的行為表現。病人是從相信自己有關節炎的部分，以為關節炎擴散到他的大腿上了，如果以這種方式重新詮釋，他的擔憂就實現了；當醫生告知他「正確」的觀念時，他也不會放棄這樣的擔憂。

第四種方式，是將 P 的錯誤解釋成一種後設語言式(metalinguistic)的錯誤。例如，將 P 的信念詮釋為：其實 P 是相信說他大腿上的病症可適用中文的「關節炎」一字。這種方式不會否定 P 仍具有一個對象層次的信念，所以三、四兩種方法可以合併使用。但博之認為這樣的方式也有幾個問題。(1).當我們碰到一個對某些觀念一知半解的人時，我們不會決定所有歸屬給他的信念內容，在形式上必須是後設語言式的。(2).如果接受這種方法，那我們就不能說一個中文說話者和一個外國說話者共享了同一個關節炎觀念，因為外國人並不用「關節炎」來表達他對關節炎的想法。(3).博之認為當我們有一個態度時，我們通常也有相干的其他態度。按照第四種方法，P 的推理會被這樣描述：他相信「關節炎」一詞適用於他身上的一種病，他相信「關節炎」適用於他手腕和腳踝上的病症，而他害怕被稱作「關節炎」病症會發生在他的大腿上。問題是說，很難根據證據而把相應於這些後設語言的對象層次信念歸屬給 P。如果有人堅持心理事件真的含有一個後設語言式的內容，這也會是一種特設假說。

博之另一方面也指出，在某些情況中，我們的確不會以字面意義的詮釋策略來理解說話者的話語，而會根據其他的標準來重新詮釋主體的心理狀態(IM, pp.42-46)。比方說外國人或小孩子剛開始學詮釋者的語言的情形，或者說話者說的是地區性方言，或者說話者相當徹底的誤解某些觀念的時候（例如，把「猩猩」當作一種飲料的名稱）。而還有一種情形是如果沒有進一步說明，就會產生誤導的情況：例如我知道 P 誤以為要有正式文件才叫定契約，而 Q 和 P 訂了契約卻

²⁰ 這是說，假設真有「大腿炎」這樣的詞彙，由於「大腿炎」僅僅應用到病人 P 所感覺到的大腿病痛，病人 P 給不出一個客觀的限制，說明「大腿炎」的用法在什麼情況下算是正確的、什麼情況下是錯誤的。所以，我們無法決定所謂的「大腿炎」到底指稱什麼。

沒有寫下正式文件。當 Q 問我說 P 怎麼看待這個約定的時候，我光和 Q 說：「P 以為你沒和他定契約」或者「P 認為你沒有什麼法律義務」，就會是一種誤導。在這種情況中，我有責任去說明 P 的其他信念：「不過他以為契約都要有正式文件才算」。博之認為，這種情況「(...)說明了在報導單一態度內容時，我們一般主張主體具有一系列的其他相干態度。某些其他態度提供了此報導的理由。在這些情況中，記錄並闡明主體異常行為的本質與範圍是很重要的。否則預測與評估該主體的行為與思想，就會有所偏差。當異常行為非常異常時，歸屬會要求重新詮釋該主體的話語。徹底誤解或者心理不穩定，會是這種情況。不過，一般的踐行似乎都允許我們，透過闡明主體的變異行為來抵銷誤導的主張，在我們那時的心理歸屬中保持字面意義地詮釋他的話語」(IM, p.45)。

不過，博之並不認為他的思想實驗的第一步，屬於上述的這些需要重新詮釋或者補充說明的情況。

第三節 博之的可能辯護



根據博之的這些說法，我們可以看出戴維森的批評並不合理。在戴維森所設想的例子以及他的詮釋中，他認為說話者 S 的關節炎信念部分是由「其他大量而含混的假設」所決定的，如果 S 的某些相干信念被證明是假的，S 就會改變他報導這個信念的語詞用法（「關節炎」）。從前面的分析我們得出說，S 在未被其他人糾正之前，他用「關節炎」一詞所指的是他心目中認定的鈣質關節炎，而在他被糾正之後，他用「關節炎」一詞所指的是醫生或一般人所認定的關節炎。在兩個情況中，似乎我們都不能說 S 的用法有錯。

然而，在這樣的詮釋下，S 事實上是有兩個關節炎概念，而不是一個。這樣的詮釋有著博之所反駁的第三種重新詮釋策略——歸屬對象層次信念——的問題。如果 S 原先使用「關節炎」的用法，不能被說成是錯誤的，那麼即便醫生或其他人告訴他「關節炎」可用在不因為鈣質堆積在關節處所引起的病痛，S 也沒有理由改變他的用法——因為他們根本意指不一樣的東西。當然，我們或可回應說，S 的信念之中或許包含著一個信念是，我所謂「關節炎」就是指專家或一般人以「關

節炎」所指稱的那種病症；所以，當他受到專家或一般人的指正時，他會修正自己的信念，改變「關節炎」的用法。不過這種回應的方式，恰好是博之所反駁的第四種方法：以後設語意學的方式進行重新詮釋。這種方式不僅是一個特設假說，有著博之所批評過的問題，也讓我們有理由說，S 將專家或一般人的用語標準當作自己用詞的正確標準，並以這個標準來認定 S 之前的用法是錯誤的。

實際上，從博之對合約例子的分析以及上一節的引文來看，他也注意到了思想者的其他信念在心理歸屬中所扮演的重要角色。然而，他的主張是，我們在心理歸屬中補充說話者其他的信念，用意是在保持以字面意義的方式來詮釋說話者的話語，以抵銷可能誤導，而非以說話者意義重新詮釋說話者的話語。如果我們是以說話者意義來重新詮釋說話者的話語，那麼戴維森所要求的補充說明就是不恰當的。在第一節所討論的反例中，戴維森主張，在 Q 對 Y 的所轉述的心理報導中，Q 必須補充說「S 與 R 都相信你有關節炎，只是 S 認為關節炎一定是由鈣質堆積引起的」。然而如果 S 使用「關節炎」的方式不能說是錯的，那麼 S 用「關節炎」一詞事實上所指的東西和 R 就是不一樣的。在這樣的情形下，Q 的報導必須修正如下，才是恰當的報導：「R 相信你有關節炎，而 S 相信你有他所謂的關節炎。S 所謂的關節炎是指一定是由鈣質堆積所引起的那種關節病症」。

目前的情況是，如果在戴維森的心目中，S 和其他人用「關節炎」想要指的是同樣的東西，那麼他沒有理由說我們能夠以說話者意義的方式來重新詮釋 S 的「關節炎」一語。如果上述的論證是合理的，那麼我們可以說 S 的「關節炎」一詞的意義和其他人沒有兩樣，只是 S 使用「關節炎」一語的方式，顯示出他對關節炎有所誤解或一知半解而已。這再一次顯示出社會外在論的合理性：相對於作為社群用字標準的其他人來說，一般的說話者在完全掌握一個語詞的正確用法之前，就已經在使用這個語詞意指某物、進行溝通，只要說話者能表現出一定的語言能力，以及（隱然）導向社群標準的用字調整，那麼我們就能根據社群的標準，字面意義地詮釋說話者的話語，歸屬他相應的心理內容。

當然，戴維森或許只是想要表明，一個說話者在日常的溝通活動中並不總是按照著社群的標準來說話；只要說話者能以詮釋者能夠瞭解的方式讓詮釋者理解他的話語意指為何，那麼就算說話者使用語詞的方式偏離了社群常規，我們仍然能夠以說話者的意圖以及其他的相干信念，決定出他的話語表達了什麼樣的心理

內容。戴維森認為，雖然在實際的溝通活動中，有著社會性的組織（例如法院）在規制著人們的用語方式，但這不代表一個人的語言意義一定被限制在社群的標準上。(KOM, p.27)如果戴維森的反對意見是這樣，那麼問題就會回到博之的社會外在論僅僅是描述性的理論、還是規範性的理論。社會外在論如果只是描繪一般說話者的語言現象和用語習慣，那麼戴維森所提出詮釋模型顯示出語言活動中的另一個面向。兩者基本上是可以相容的。但如果社會外在論是一套規範性的理論，主張社群的常規總是詮釋一個說話者之話語的標準，那麼戴維森的詮釋模型就會與社會外在論有所衝突²¹。

筆者相信博之的社會外在論應是前者而非後者，理由在於博之談到許多要以說話者的主觀條件進行重新詮釋的例子（見第二節）。博之也明確指出說話者能夠有意識地退出語言常規、形成自己一套語言用法：「在親屬社群中出生、成長的一個人，或許會單純地（單方面）決定他要遵從地區方言的用法，或甚至形成他自己對特定語詞的用法，有意識地退出親屬社群在這些特定語詞上立下的常規。在這樣的情況裡，親屬社群的成員不會，也不應該，根據對他的語詞同音異義的解釋來歸屬他心理內容。在此，個體對共有常規與共有概念的意向或態度，似乎比他處理一個字的因果前件來得重要——除非那些意圖單純地被包括在追本溯源的故事裡。」(IM, p.78)在這個意義下，我們可以說戴維森與博之兩人外在論並沒有本質性的衝突之處；兩人只是在強調語言現象中兩個不同的面向。

戴維森不信任社會外在論的其他兩個理由，在筆者看來也不是反對社會外在論好理由。戴維森第一個理由是，他認為以說話者的意圖來詮釋他的話語，比起我們以其他人的用語標準來詮釋說話者的話語，要來的恰當。他似乎是認為，依其他人的用語標準來進行詮釋，就會忽略了說話者使用語詞的意圖。我們必須指出，按照博之的說法，我們能以字面意義的方式詮釋說話者的話語的條件在於，說話者必須要能先表現出一定的語言能力與對社群標準的責任感(responsibility)。博之對說話者的責任感的強調，就是以說話者的意圖來說明，為什麼我們能以社群標準來決定說話者的用字標準：因為說話者隱然地企圖以其他人的用語標準，作為自己的用語標準；說話者修正自己語詞的用法，正好表現這

²¹ 這一個問題是由口試委員方萬全老師所指出，是一個值得繼續深究的核心問題。

點。

戴維森不信任社會外在論的第二個理由在於，他不信任博之和普特南所使用的思想實驗，能夠顯示出人們在實際溝通活動中表現。然而，我們必須指出戴維森自己也使用思想實驗進行論證，例如他以沼澤人(swampman)的思想實驗，顯示出因果歷史是決定一個人心理內容的因素之一(KOM, p.19)，並以此反駁普特南的雙生地球思想實驗(KOM, pp.32-33)。另一方面，筆者認為這不是個好理由的原因在於，筆者認為思想實驗的重點不在於這些假設性的事件是否可能實現，而在於我們能夠透過思想實驗給我們的直覺反應，顯示出真正與議題相干的概念為何。雙生地球思想實驗在直覺上告訴我們，隨著物理環境與其他人的轉變，我們的心理內容和語詞意義也會跟著改變；它顯示出了，我們在設想如何決定心理內容和語詞意義時，什麼事情才是決定性的因素。

第四節 博之論自我知識

在討論過戴維森的批評後，我們澄清了博之的外在論容易引起的問題。接下來我們要討論博之對自我知識的看法為何，並說明他是如何解決外在論與自我知識表面上的不相容論。

博之認為，自我知識指的是能構成知識的自我歸屬判斷，例如「我現在正在思考(I am thinking now)」、「我判斷\懷疑說(I judge\doubt that)水比水銀常見」或者「我認為說(I think that)寫作需要專注」。這些判斷能成為知識，並非透過經驗調查而被得知，而以一种特殊的方式自我證實的：下判斷本身就使得此判斷為真。在這個意義下，我們對自己的心理狀態與內容具有直接而權威性的自我知識。博之稱這樣的判斷為「基本自我知識」(basic self-knowledge)²²。

在此，「基本」一詞是指，相對於像我相信說____、我欲求說____這類的自我知識，我想說____、我判斷說____這類我思般(cogito-like)的判斷具有特殊的認

²² Burge, T.(1988), 'Individualism and Self-Knowledge', in *Externalism and Self-Knowledge*, ed. by Peter Ludlow and Norah Martin. Stanford, California: CSLI Publications, 1998, p.111-112. 以下簡稱此文為'ISK'。

知地位，是不可錯的、脈絡上自我證實的判斷²³。基本自我知識類似於沒有東西是狗又不是狗這樣的簡單邏輯真理，由於其形式與邏輯的關係，一定是真的。在認知上，不需要藉由經驗或推論，一個思考者憑著理解這樣的命題就足以知道此命題真。不過基本自我知識與簡單邏輯真理的差別在於，這些判斷要被思考才會是真的，並且是以一種特殊而直接的方式自我證成的²⁴。

根據博之的內容外在論，我們的心理內容部分是由思想者與（物理或社會）環境之間的關係所決定的。然而，從我們可以對自己的心理內容有直接的、非經驗性的知識來說，這是否意味著我們的權威性自我知識是獨立於環境而獲得的呢？博之指出，笛卡兒所提出的個體論模型，暗示了我們可以獨立於環境、僅僅憑藉著內省而獲得關於我們心理內容的知識(ISK, p.113)。博之提出一個論證來反駁這樣的主張。如果我們能憑藉著內省而個別化自己的思想內容（知道自己在想什麼），那麼在一個人在不同的環境中具有不同的思想，他必定能透過內省在不同的情況中知道兩思想為何，並知道兩思想在不同的情況中是不同的思想。但是我們可以設想一個稍做修改的雙生地球思想實驗，顯示出一個人的思想隨環境而改變了，但他卻無法以內省的方式察覺到這點。

假設存在一個雙生地球，上面具有與地球幾乎完全相同的對應物(counterpart)，只除了雙生地球人稱之為「水」的東西具有完全不同於地球水的分子結構 XYZ。再假設實際地球上的一個人 O，他在睡夢中被轉換到了雙生地球上，一去就是幾十年。由於雙生地球與地球表面上完全相同，我們可以肯定 O 並不會察覺到有什麼不同之處而以為自己住在地球上。假設在十幾年之後，有人告訴 O 說他是從地球上被轉換到雙生地球上的，並問他說知不知道是哪時候來的，O 能否光靠內省來知道這點呢？直覺上，O 顯然辦不到這一點。這顯示出，

²³ Burge, T. (1996), 'Our Entitlement to Self-Knowledge', in *Externalism and Self-Knowledge*, ed. by Peter Ludlow and Norah Martin. Stanford, California: CSLI Publications, 1998, p.241. 博之認為，由我思般的判斷所代表的基本自我知識，是說明權威性自我知識的特殊認知地位與不可錯之特徵的範例。與此相較，我們對其他傾向性心理狀態的自我歸屬，像是第一章提到的「我相信說_____」、「我欲求說_____」等語句所表達的判斷，沒有不可錯、脈絡上自我證成這樣的特徵。這並不表示說，一般的自我知識觀念並不包括關於傾向性心理狀態的自我歸屬。對博之來說，因為其他的理由，在一般的情形中，我們一定知道自己相信什麼、欲求什麼。我們會在第五章討論這個理由。

²⁴ *Ibid.*, p.242. 邏輯真理的真值由其邏輯形式就可顯示出來，是否有任何特定的思考者設想過這樣的真理，不影響這樣的命題為真。我思般的判斷雖在『光憑理解就能知道此命題為真』形式上類似於邏輯真理，但是這樣的判斷的真值條件包括被一個思考者思考。我思般的判斷唯有被一思考者實際的設想時，才會為真。

O 在歷經環境的轉變後擁有了不同的思想（好比說關於水和雙生水的思想），而他無法光憑著內省來知道環境的改變，以此分辨兩種思想的不同(ISK, p.115)。

需要注意的是，這個思想實驗並不是說 O 因此不知道自己的思想為何，也不是說 O 需要調查一下兩地球的還境差異才能知道自己的思想為何。這個思想實驗要顯示出的是，光靠一個人對自己心理內容的省察，是無法個別化自己的心理內容為何的。可是，問題就在於，如果我們的思想內容隨著環境改變了，而我們並不能透過內省來分辨它們，那麼為什麼我們不需要透過調查環境的方式來辨別它們，我們就有理由說自己知道自己在想什麼？

博之指出，這是由於思考特定思想的條件，在我們進行思考的時候已經預設在其中，而我們不需要知道這些複雜的條件是否已經成立，才能進行思考。按照物理外在論，要對某物進行思考，必定要和某物發生因果關係。好比說要思考水是液體這樣的內容，必定要和水發生（直接或間接的）因果關係。要知道這樣的因果關係是否成立，一個人必須透過經驗的調查才能得知。不過，在我們思考水是液體時，我們不需要知道這些複雜的成立條件(enabling conditions)；這些條件只需要被預設即可(ISK, p. 116)。然而，為什麼我們不知道這些條件是否成立，我們就有理由說我們是在思考水是液體呢？博之以知覺知識為類比，來說明這點。

當我們說自己具有什麼樣的知覺知識時，我們不事先知道某些相干的成立條件已經滿足了，才有理由這麼說。比方說當一個人知覺到角落裡有一塊錢的時候，我們不需要此人事先知道讓他的知覺信念成為可能的那些因果條件，比方說他的眼睛功能是否良好、光線是否充足、他看到的是鏡相還是實物，以及他是否能分辨角落裡的那一塊錢是真幣還是偽幣...等。在不假設世界知識的懷疑論的前提下，我們可以肯定說，只要這些條件都成立了，思想者就有理由形成我看到角落裡有一塊錢的知覺知識(ISK, p.117)。類似的，當我意識到水是液體時，我同樣不需要事先知道使這樣的思想成為可能的成立條件，才有理由形成這樣的思想；只要相干的條件被預設了，我就能夠設想這樣的思想。而當我判斷說我認為水是液體時，這個判斷需要我在同一個思考過程中思考水是液體這個一階思想，因此也預設了使得我能夠思考這個一階思想的成立條件。同樣的，要擁有這樣的二階的思想，思考者不需要事先知道這些成立條件為何。

不過我們可以繼續追問：以知覺信念的類比來說，我們的知覺信念並不總是為真，在知覺條件良好的情況下，知覺可能會產生純粹的錯誤(brute error)²⁵，可是為什麼基本自我知識在下判斷的時候一定為真呢？博之指出，知覺信念與自我知識的判斷在兩方面是不同的。第一，知覺信念產生純粹錯誤的可能，來自於『知覺對象獨立於知覺者的認知機制』的事實。知覺對象與知覺者的關係並非必然的：我們在任何特定的場合中，保持同樣的知覺機制，由於我們的知覺無法在表面上區分知覺的對象與幻覺（或相干的贗品），我們的知覺判斷總是有純粹錯誤的可能(ISK, p.120)。但基本自我知識不會有純粹錯誤的可能，因為在自我知識的情況中，判斷與判斷之主題的關係並非偶然的。這些思想是自我指涉(self-referential)與自我證實的。以我判斷說：我正在想說水是濕的為例。此判斷是自我指涉的是指，當我做出這樣的判斷時，我正在判斷的認知內容，由這個判斷本身所確定出來。自我證成指的是，我做出這個判斷的動作本身，滿足了此判斷的真值條件(Ibid.)。這樣的情況保證了，在我的相干的認知機制保持正常，而我做出我想說_____這類的判斷時，不可能出現純粹錯誤的情況（也就是說當我們自我歸屬說「我在想 p」的時候，不可能出現我實際上在想 q、但我卻誤以為我在想 p 的純粹錯誤），而且這個判斷總是為真。

第二，知覺知識是客觀的、非個人的(impersonal)。這是說任何人處於相同的認知立場與角度，在相干的時間點上都能對同樣的知覺對象夠做出同樣的觀察。(ISK, p.120)例如一個人在某一時間點做出的知覺判斷：楓林小徑兩旁有許多綠樹，其他人若在不同但相近的時間點站在同樣的觀察角度，也能做出同樣的知覺判斷。知覺知識的認知保證不受到知覺者的觀點與判斷時間的影響。但是自我知識本質上是主觀的、個人式的；基本自我知識的特殊認知資格，依賴於『下判斷』與『判斷的內容』兩者都出自於同樣的觀點與發生時間。(Ibid.)以上述的例子來說，當我判斷：我正想說楓林小徑兩旁有許多綠樹時，我出於第一人稱觀點做出這樣的判斷，而此判斷內容中的我指稱同樣的第一人稱觀點；而我做出這個判斷的時間與被判斷的思想發生的時間，是同樣的。唯有如此，這個判斷才能是脈絡

²⁵ 純粹錯誤是指，在主體的知覺機制良好，環境條件配合的情況下，由於幻象或知覺對象外表上的相似所產生的錯誤判斷。例如，在我視覺正常的情況下，我看到一顆蠟做的蘋果，而誤以為那是一顆蘋果。

上自我證實的基本自我知識。

為什麼在理解外在論與自我知識時，會產生不相容的表面問題呢？博之診斷說，因為，我們在設想雙生地球思想實驗的時候，轉換了一個第三人稱的觀點，以一種外部的角度來注視我們自己。在這樣的觀點中，我們很自然會設想說，除非我們能夠完全掌握相干的外在環境因素，否則我們就沒理由說我們『知道』自己在想水還是雙生水。然而這樣的自我知識觀念，是一個扭曲的觀念(ISK, p.124)。在雙生地球的轉換情況中，一個人的一階思想隨著環境的轉變而發生了變化，但是他卻不需要調查環境、比對環境之間的差異，才有理由說他知道自己正在想什麼樣的思想。一個人的自我知識的認知地位，僅僅來自於此人在反思性的思考中，以某個思想示例(thought tokens)作為某思想示例的方式來個別化一階的思想，並將之歸屬給自己。這個人不知道環境的轉變、無法以此分辨不同的思想，無關乎這個二階判斷是否為真、是否被證成。

博之指出，笛卡兒的錯誤在於，將自我知識混同為一種知覺知識，主張一個人可以掌握自我知識所有的成立條件，使得自我知識比知覺知識還要確定(ISK, p.122)。這會讓人以為自我知識的主題——思想者的思想——就像知覺知識的對象一樣，需要透過某些調查才能得知。在這樣的情況下，如果一個人的思想內容要取決於外在事物的話，他就不能直接掌握自己的心理內容，具有非推論性的自我知識。然而，思想實驗指出，笛卡兒式的內省對個別化一個人的思想來說並非決定因素，對一個人的自我知識來說更不必要。博之認為：

…我能知道我有關節炎，並且知道我認為我有關節炎，即便我沒有恰當的判斷說何為關節炎。要設想一個人的思想、並因此設想我思般(cogito-like)的思想，一個人必須充分理解一個人在想什麼才能去設想它，是一個自明之理。不過這個道理並不會導出說，這樣的理解帶有一種能力，能藉由其他的思想與概念，去正確地解釋一個人的思想或概念；也導不出說，這種理解不會有解釋上的錯誤。所以一個人能夠知道一個人的思想為何，即便他對自己的思想或概念做出了不完全的或者錯誤的說明。一個人不應該將基本自我知識意義下的「知道自己的思想為何」，比做為能正確解釋這些思想——能夠描繪這些思想與其他思想的構成關係——這個意義下的「知道自己的思想為何」。(ISK, p.125)