國立臺灣大學

哲學論評

NTU PHILOSOPHICAL REVIEW 第四十七期

目 次

論　　著

ARTICLES

王植對《注解正蒙》神化觀之批判
－以「太虛」三層義為進路

陳政揚[*]

摘　要

　　設若研讀哲學家的代表著作，是吾人探究其思想的必要條件之一，則逐句解析張載思想論述的明清《正蒙》注，顯然是研究橫渠思想在明清哲學之發展時，所無可忽視的一環。尤其歷來注解《正蒙》者，不乏高攀龍，王夫之等望重士林，或在哲學史上具有獨創見解者。然而在當代張載學研究中，相關議題似乎仍較少引起學者間的討論興趣。基於此，本文以清代《正蒙》學研究為論述起點，旨在澄清李光地在《注解正蒙》中對張載神化觀的詮釋，以及王植在《正蒙初義》中如何由釐清「太虛」概念，批判李《注》之非。全文共分為四項研究環節：首先，由於王植以澄清「太虛」概念為解讀《正蒙》與衡定諸注之關鍵，本文先扼要說明「太虛」之三層義。其次，本文將指出李《注》對「太虛」之詮釋，以及王植對李《注》太虛義之批判。再者，由於李《注》不僅嘗試解消「太虛」之本體義，而且以理氣二分的理論架構，詮釋橫渠虛氣一體之神化觀。本文將指出李《注》對橫渠神化觀之新詮。最後，本文將藉太虛三層義，指出《正蒙初義》的洞見以及李《注》神化觀的侷限。

---

[*] 陳政揚，高雄師範大學經學研究所副教授。

A Study of Wang Zhi's Criticisms
of the Deification
in the *Commentary of the Zhengmeng*
by an Analysis of the Three Aspects
in the Idea of "Ultimate Voidness (*Taixu* 太虛)"

Cheng-Yang Chen[*]

Abstract

In order to understand the content and significance of Zhan Zai' thought, commentarial works on the *Zhengmeng* during the Ming and Qing period are indispensable because these works present detailed explications on every conceptual terms. Compared to those commentarial works on the Zhang Zai's thought by Gao Panlong and Wang Fuzhi who have been regarded as rather creative thinkers, other commentarial works during the Qing period have not drawn sufficient scholarly attention. Bearing this in mind, this article aims to explore the two topics: Firstly, it analyzes Li Guangdi's interpretation on Zhang Zai's idea of "deification" through Li's *Commentary of the Zhengmeng*. And secondly, it also traces Wang Zhi's elucidation of the concept of "Ultimate Voidness" (*Taixu*) in his commentarial work, the *Basic*

---

[*] Associate Professor, Institute of Confucian Classics, National Kaohsiung Normal University.

Meanings of the Zhengmeng, as well as Wang's correction of initial mistakes in aforementioned Li Guangdi's commentary.

The discussion of this article consists of the following four sub-topics: (1) Wang Zhi's exposition of three aspects in the concept of "Ultimate Voidness"; (2) Li Guangdi's interpretation of the term "Ultimate Voidness" and Wang Zhi's repudiation against it; (3) Li's argument that the ontological significance of the concept of "Ulitmate Voidness" should be dissolved, and that Zhang Zai's idea of deification of continuity between vital breath and voidness should be understood by the dualistic framework between principle and vital breath; and (4) Highlighting of Wang's insight and Li's limitation of his conception of deification by means of our analysis of the three aspects in the concept of "Ultimate Voidness."

王植對《注解正蒙》神化觀之批判
－以「太虛」三層義為進路*

陳政揚

壹、前言

　　在明清《正蒙》注解發展史中，李光地（榕村，1642-1718）的《注解正蒙》是清代《正蒙》注釋中最具代表性的著作之一。《四庫全書提要》曾指出李注的學術貢獻，不僅在於針對先儒「不敢置議」處提問探析，而且又對於先儒互異之處，能「皆一一別白是非，使讀者曉然不疑」，甚至闡發張載未發之意，故以為本書於「明初以來諸家注釋之中，可謂善本」。[1]又由於李氏在政治與學術上的影響力，其注釋也廣泛影響了其後的《正蒙》詮釋者。[2]因此當清儒王植（戁思，1682-1767）全面梳理歷

---

[1] 《欽定四庫全書‧注解正蒙‧提要》曰：「《正蒙》一書，張子以精思而成，故義博詞奧，注者多不得其涯涘。又章句既繁，不免偶有出入，或與程朱之說相牴牾。注者亦莫知所從，不敢置議。光地是書疏通證明，多闡張子未發之意」。關於《注解正蒙》的版本源流，可參考胡元玲、林樂昌先生的研究成果。本文引用《注解正蒙》皆依據《欽定四庫全書》，凡衍誤的字句以圓括號「（　）」標示，而補脫與改正之字句以方括號「〔　〕」標示。詳見（清）李光地，《注解正蒙》，頁 1-2；胡元玲（2004：247-248）；林樂昌（2012）。

[2] 李氏深受清康熙帝的信任，在當時政治學術上深具影響性，康熙曾言：「知光地者莫若朕，知朕者莫若光地」。當代不少學者亦曾深入探討李氏之政治背景，及其理學與經學之關聯。

來《正蒙》諸注時，他即以李注為主要論辨對手與對比主軸。在《正蒙初義》中，王植徵引李注共 119 次。其中純粹徵引李注與其他《正蒙》注解並列參考者有之。例如，在注解〈動物〉之「人之有息，蓋剛柔相摩、乾坤闔闢之像也」，王植即將明儒徐必達（德夫，1562-1631）的《正蒙補訓》與李注並列，提供讀者參考。其中，肯定李注者有之。例如，在注解〈參兩〉之「雷霆感動雖速，然其所由來亦漸爾。能窮神化所從來，德之盛者與！」時，對於《初義》所收錄的《正蒙》諸注中，王植便肯定的表示：「惟李注得之」。但在《正蒙初義》中，王植更多的是扮演著批判者的角色，逐一檢視李注之得失。最鮮明的例子之一，就是李注對〈參兩〉篇首之言著力甚多，他不僅批判張載對〈說卦〉「參天兩地」說的解釋未確，更表示橫渠此論實為啟發胡宏（五峰，1105-1161）「無對之善及與惡對之善」之說的源頭，陷入朱子所謂「三角底太極」之嫌。[3]王植卻認為這僅是由於張載取義不夠直捷，而「不免添出一層而分配處亦未確」。[4]換言之，《正蒙初義》並非僅是以《正蒙》注解集成的形式收錄李注。或可說，自李光地完成《注解正蒙》後，王植是第一本，也是至今最為完整的一本全面檢視反思李注的著作。故本文認為，若吾人欲探究清代張載學之發展，則李注與王植的評斷，顯然是無可忽視的資料與題材。

　　至於選定神化觀為探究王李二人注解異同，乃由於自朱熹（晦翁，1130-1200）表示：「『神化』二字，雖程子說得亦不甚分明，惟是橫渠推

本文不再贅述。詳見（清）章梫纂，〈任賢〉，《康熙政要》，卷 4；史革新（2007：186）；徐世昌（2010：1531-1623）；楊向奎（1985：690）；姜廣輝（2010：124-149）。

[3] 在《注解正蒙》中，李光地雖有意調和朱、張之學，但吾人不可忽略李氏的尊朱立場。在《榕村語錄》中，他即曰：「周、程、張、邵不得他（朱子），恐不能如此烜赫」。（清）李光地，《榕村語錄》，卷 19，頁 333。

[4] 由於《正蒙初義》除在卷首收入〈臆說〉作為解析全書之總綱外，還依照《正蒙》十七章分為十七卷，每卷頁碼以「一」起始。因此，下文引用〈臆說〉將直接標示頁碼，而引用《正蒙初義》時，則標示「卷」與「頁碼」。（清）王植，〈臆說〉，《正蒙初義》，頁 3。

出來」[5]之後，歷代《正蒙》的詮釋者，都留意到「神化」這組概念在橫渠氣論中的重要地位。[6]例如，王夫之（船山，1619-1692）在闡釋〈神化〉通篇要旨時，即曰：「此篇備言神化，而歸其存神敦化之本義，上達無窮而下學有實。張子之學所以別於異端而為學者之所宜守，蓋與孟子相發明焉」。[7]又由於張載闡發「神化」觀時，是藉由虛氣論的論述架構，將《周易・繫辭下傳》所謂：「精義入神」、「窮神知化，德之盛也」，《中庸・30章》之「大德敦化」，以及《孟子・盡心上》之「君子所過者化，所存者神，上下與天地同流」相互串言，皆統納為天道性命相貫通之學。因此，《正蒙》的詮釋者面對課題有二：一者，是否接受張載以氣論嫁接不同儒家經典的詮釋方式？[8]二者，設若張載所論與所引經典已有出入，則《正蒙》所欲闡述的義理究竟為何？又何以能上承孔孟儒學之義理脈絡？李光地的注釋，與王植對李注的批判，亦是在此背景中展開的。

　　基於上述，本文共分為四項研究環節：首先，由於王植以澄清「太虛」概念為解讀《正蒙》與衡定諸注之關鍵，本文先扼要說明「太虛」之三層義。其次，本文將指出李注對「太虛」之詮釋，以及王植對李注太虛義之批判。再者，由於李注不僅嘗試解消「太虛」之本體義，而且以「理／氣」論之架構，詮釋橫渠虛氣一體之神化觀。本文將指出李注對橫渠神化觀之新詮。最後，本文將藉太虛三層義，指出《正蒙初義》的洞見以及李注神化觀的侷限。以下即依此展開討論。

---

[5]　（宋）朱熹，《朱子語類・張子書一》，卷 98，頁 3303。

[6]　誠如劉述先先生所指出，橫渠在宋代理學中的地位，實與朱子的詮釋與評價有著緊密的牽涉。劉述先（2012：1-20）。

[7]　（明）王夫之，《張子正蒙注》，收入《船山全書》第 12 冊，頁 76。

[8]　例如，王植雖認為張載是「引用多與本旨弗類」，但仍肯定他的詮釋不離儒家經典的義理，故表示「而經傳所言神化之理大都具是矣」。《正蒙初義》，卷 4，頁 1-2。

貳、王植論《正蒙》之「太虛」三層義

　　誠如《正蒙初義・提要》撰者所舉，王植綜觀歷來《正蒙》諸注之爭端，多源於對橫渠所論「太虛」之義不明。[9]故不僅表示：「『太虛』是張子主見」，[10]嘗試釐清《正蒙》所論之「太虛」共有三層義，更依此辨析南宋至清初《正蒙》注釋之爭端。因此，本節先扼要陳述王植論《正蒙》「太虛」之三層義。在下節中，則依據此太虛義探究王植對《注解正蒙》之批判。

一、分論太虛之各層義

　　在《正蒙》中，「太虛」實包含三層相互串連的意義。對此，在《正蒙初義・卷一》中，王植指出：「太虛有以未發之體言者，有以流行之用言者，有以究極之歸言者」。基於《正蒙》各篇章文脈之不同，「太虛」概念亦可析解成以下三種意涵：

　　其一，就太虛為未發之體而言，太虛的第一層義，即是作為生化萬有之本體。依此，太虛具有整全性（「渾然未形」）、先在性（「為天地萬物之大母」），以及根源性（「在造化則本此以生天地」）等性徵。[11]

　　其二，就太虛為流行之用而言，太虛之第二層義在於其發用流行即是造化之妙運萬物。在《正蒙》中，張載是以陰陽二氣之聚散浮沉、升降飛揚，闡述萬物如何在無止息的氣化絪縕中得以生生不息。他雖說萬物形色本諸陰陽氣化，陰陽氣化本諸太虛，但尚未明言陰陽氣化即是太

9　（清）紀昀／等，〈提要〉，《正蒙初義》，頁 1-2。
10　〈臆說〉，《正蒙初義》，頁 4。
11　〈臆說〉，《正蒙初義》，頁 2。

虛。但由《正蒙初義·卷一》曰:「凡二氣之絪縕,天地之對待,陰陽之流行皆太虛之妙用」可知,王植已然由太虛之第二層義表明「太虛」與「氣化」並非一者為「體」、一者為「用」,體、用分判為「二」的關係,太虛本就是即體即用之造化自身。再者,若說在太虛第一層義中,所呈現的是太虛之整全性、先在性,與根源性。那麼在第二層義中,他不僅通過二氣絪縕以生萬物而指出太虛的生物性,亦藉由氣化流行週遍天地萬物而呈顯出太虛的遍在性。

其三,就太虛為究極之歸而言,太虛之第三層義,即在於闡明萬物之死生流轉不外乎是一氣之聚散流行,並由此反駁釋氏輪迴說的論點。[12]例如,在《正蒙初義·卷一》中,王植藉由注解「聚亦吾體,散亦吾體,知死之不亡者,可與言性矣」時,即表示「『散亦吾體』乃太虛第三層正義也」。[13]他認為,張載既然以太虛作為一切存有之根本,又以太虛之發用乃一無止息的氣化流行;那麼,在理論上顯然亦必須通過「太虛」說明萬有死亡後該當歸於何處,而此正是太虛第三層義所處理的課題。

二、綜論太虛三層義之貫穿

王植雖表示太虛之三層義:「其立言不一,皆當隨文體認細玩篇中之意」。[14]但他也不忘提醒讀者:「各節所言之太虛,言雖殊而意自貫,左

---

[12] 〈臆說〉曰:「至萬物散而為太虛,散亦吾體,此又從既生人物之後而要其終。先儒謂其流乃是大輪迴者,此也。此第三層」(頁 2)。又由「客形變化之小聚散,非萬物復聚為太虛之大聚散也」(頁 30)可知,他將張載以氣之聚散解釋萬物之死生流轉的說法進一步發揮,而為:1.「小聚散」是指個體隨氣聚而生、氣散而死,並在生存歷程中隨氣化活動形形相繼。舉例而言,人由氣聚而生,死時則身體形貌亦隨氣散而毀朽。當人生存時,儘管各年齡期(例如,五歲、十五歲、五十五歲)的形貌均不相同,但是此人仍屬於「同一個體」。王植以為,這是由於個體隨著氣之聚散方能在不同時空中持續維持著形貌相繼相續,而這都屬於客形變化之小聚散。2.「大聚散」則是就天地萬物最終都會在氣化活動中復歸為太虛而言。

[13] 〈太和〉,《正蒙初義》,卷 1,頁 17。

[14] 〈太和〉,《正蒙初義》,卷 1,頁 13。

右逢原而各盡其致矣」。[15]在總論〈太和〉一篇之要旨時，王植曾以「生生之原」、「生生之具」，以及「形潰反原」串言太虛之三層義：

其一，就太虛第一、二層義之串言，太虛既為生生之本體，也即是生生之氣化妙用。「生生之原」與「生生之具」並非「形上之理」與「形下之氣」的截斷關係。王植有別於前人之處，在於他既接受程朱對張載論道體未能盡明的批判；也順此指出，正由於《正蒙》所論道體尚未至程朱所見之理體，故歷來注者之誤，正是誤將張子所言太虛視同朱子所見之形上之理。[16]依此，他認為在張載思想中，太虛並非高懸氣化生物活動之外的形上本體。太虛即為湛然無形之氣，為「形」所拘限之萬物，其存在皆本於「無形」而足以「形形」的太虛。太虛不僅是貫穿有形無形的生物本體，也即涵蘊氣化活動。故太虛之一二層義串言，乃是標示造化生生之本源與妙運，本同出於太虛。注〈乾稱〉：「凡可狀皆有也，凡有皆象也，凡象皆氣也」一節時，他即指出此三句雖「只以一氣字分析」，但細究當可發現此實是「以（太虛）第一層與第二層串言」（頁39）。

其二，就太虛第二、三層義之串言，太虛既是萬物存在之場域，也是萬物死後形體消散之最終歸趨。若說一二層義串言在於澄清生生之「體」與「用」皆源自於太虛，二三層義串言則是聚焦於「生生之用」，旨在說明萬物之生成壞滅並非由「有」歸向於「無」，而是萬物依氣聚氣散之不同型態，恆存為太虛氣化之大「有」。依此，王植認為《正蒙》凡提及「神」之處，皆指向太虛第二層義，亦即由太虛生生之妙用，說明存在世界中的個體與現象皆是神化之跡。[17]以二三層義串言，則在彰明

---

[15] 〈臆說〉，《正蒙初義》，頁3。

[16] 例如，在解釋「由太虛有天之名」、「合虛與氣有性之名」時，王植即曰：「惟此若必以『理』言之，方可通」，但他緊接著表示：「然亦張子所見止至太虛而止，若遂以為生物之理之盡乎此也者，而實未盡也。故朱子雖亦以『理』字釋之，而亦謂其『生受辛苦』，『聖賢便不如此說』。蓋非理字不足為言，而直以理字代之則不可也」〈太和〉，《正蒙初義》，卷1，頁13。

[17] 〈太和〉，《正蒙初義》，卷1，頁13。

橫渠對佛道兩家生死觀之批判。王植以「一神兩化」呼應〈乾稱〉「存順歿寧」之旨，認為不僅萬物之生成屬於神化之妙用，萬物之死後復歸太虛，亦是造化必然之歷程。「生生之具」與「形潰反原」既非可以逆轉的生物歷程，亦非彼此割裂的兩個世界。由此他承繼張載批評道家追求久生不死，乃是「徇生執有者物而不化」，並且反駁佛教追求滅盡無餘的涅盤境界，乃陷入「彼語寂滅者往而不反」的困境，同屬不識造化生生之道的言論。因此，在詮解〈太和〉「知死之不亡者，可與言性矣」時，王植不僅肯定李注以「適得吾體，不失吾常之意」得橫渠此段要旨，更表示：「散亦吾體乃太虛第三層正義也。既以太虛立言，其歸結必至於此正其立論」，而「所謂『死之不亡』，即末篇『形潰反原，游魂為變』意。正以此明輪廻之妄」。[18]

其三，就太虛三層義之相串言，王植是將存有之根源、造化之活動，以及個體向整體之復歸等層面統攝於「太虛」一名中，此所以能既說：「太虛一而已矣」，但又可分三層相互串言。在注解〈乾稱〉之「太虛者，氣之體」時，即曰：「太虛，體也；神即太虛之用也。其『用』隨氣所感而善應，『體』則湛然而至一。亦以第一層第二層並舉也。至『形潰反原』並及第三層矣。蓋相感無窮，以氣之流行言其散無數，以氣之生物言其實則湛然而一，仍然太虛之本體而已」。[19]王植再次呼應詮解〈太和〉通篇要旨時的立場，以「無形而足以形形者，是太虛之本體」，揭露張載思想中的道體，本即是貫通形而上下、兼攝無形有形者。所謂太虛三層義相互串言，實即將存在世界之本源（「生生之原」），存在世界（「生生之具」），以及生生此存在世界之活動與歸趨（「形潰反原」），以「無形－有

---

[18] 〈太和〉，《正蒙初義》，卷1，頁17。
[19] 〈乾稱〉，《正蒙初義》，卷17，頁27。

形－形形－形潰反原」的結構，視為連續與一體的關係，而非形上理體
與形下氣化之異質異層關係。

叁、王植對《注解正蒙》「太虛」義的批判

　　李注指出，混淆周張二子所言之「太極」與「太虛」，是前人詮解橫
渠氣論時的關鍵之誤。此觀點顯然啟發王植對《正蒙》注者之批判。為
使討論明晰，以下將先扼要說明李注對《正蒙》「太虛」的詮釋，再依據
前述「太虛三層義」，探討王植對李注之批判。

一、李光地對《正蒙》「太虛」義的詮釋與開展

　　清儒李光地（榕村，1642-1718）對《正蒙》太虛義的重要洞見，在
於辨析張載所言之「太虛」與周敦頤（濂溪，1017-1073）所言之「太極」
並不相同。注〈太和〉之「太虛無形，氣之本體」一段時，他即曰：

> 言太虛無形之中，而氣之本體存焉，即太極也。朱子《圖解》云：
> 「此所謂無極而太極也」，所以動而陽，靜而陰之本體也，正此
> 意也。然周子謂之「太極」，而張子謂之「太虛」者，「太極」如
> 「性」字，「太虛」如「靜」字；「太極」如「中」字，「太虛」
> 如「未發」字。「人生而靜，天性具焉」，非以靜為性也。「喜怒
> 哀樂未發而中存焉」，非以未發為中也。太虛無形而無極之真在
> 焉，非以太虛為太極也。[20]

由引文可知有二：其一，就思想淵源上，李注指出，以「無極而太極」
詮釋「太虛無形，氣之本體」，是出自朱子《太極圖說解》的論點。但

---

[20] 《注解正蒙》，卷上，頁 1-2。

《太極圖說》中的「太極」概念，並不等同於《正蒙》中的「太虛」概念。

其二，就概念意涵上，李注舉「人生而靜，天性具焉」辨析「太極」與「太虛」之異。他指出，「太虛」如同本句之「靜」字，而「太極」如同「性」字。「靜」是使「性」得以為吾人所體察的條件，但「靜」不是「性」。同樣的，「太虛」是使「太極」得以呈顯的條件，但「太虛」並非「太極」。李注接受朱子「性即理」的論點，故當他以「太極」對比於天所賦予人之「性」時，是在本體論意義下，將本句中之「太極」與「天性」，均視為通同於形而上之「理」。但在朱子思想中，形上之理只存有而不活動，是處於「淨潔空闊底世界」。世間萬象之生成流變，均是由具活動義的氣所擔負。所以形上之理雖能宰「氣」，但就實存層而言，理卻是通過氣化流行才得以呈顯。[21] 李注亦依此架構詮解《正蒙》「太虛即氣」說，視「太極」為形上理體，但「太虛」卻僅為此「理」得以在現實世界中呈顯的無形之氣。太極與太虛之關係，可理解為「理」與「氣」的關係。

李注為凸顯「太極／太虛」具有這層「形上／形下」之別，再次以「『喜怒哀樂未發而中存焉』，非以未發為中」為例。他指出，此處之「中」並非人在「喜怒哀樂未發」時的狀態，而是指向人之「性體」，或名之「中體」。「喜怒哀樂未發」雖是使「中體」得以呈顯的狀態或條件項。但「中

---

[21] 例如，在《朱子語類‧理氣上》中，當朱子回應「先有理後有氣」之說時，先曰：「然以意度之，則疑此氣是依傍這理行。」並接著說：「若理，則只是箇淨潔空闊底世界，無形，他卻不會造作；氣則能醞釀凝聚生物也。但有此氣，則理便在其中。」。可知他雖將「理」視為「氣」所依傍的存有基礎，且認為「理」能主宰「氣」之聚散生物活動。但在理氣關係中，理是擔負主宰義的形上本體，而氣才是擔負現實義與活動義的生化萬物之載體。牟宗三先生即指出，朱子的「理」是「只存有而不活動」的「但理」（但有其理而已）；理氣關係是「不離不雜」的關係。亦即理雖不離於氣而獨存，但形而上之理與形而下之氣在存有論上有著不可相互化約的區別。《朱子全書》，第 14 冊，頁 116；牟宗三（1981：冊 3，486-516）。

體」不等同於使其得以呈顯之「條件項」。故說：「非以『未發』為『中』」。由於人之中體實根源於太極，故具有形而上的意義。藉由上述釐析，在《注解正蒙》的詮釋中，「中／未發」之別，「太極／太虛」之異，實即是「形而上／形而下」的區分。依此，本文將李注詮釋要點歸結如下：

　　（1）「太虛」是感官知覺難以察視的無形之「氣」，「無極之真」乃
　　　　　是指向形而上之太極的描述語，太極為形上之理而遍在於無形
　　　　　之氣（太虛）中。

　　（2）就實存層而言，太虛與太極相即不離（「『太虛』無形而『無極
　　　　　之真』在焉」）。但若以概念釐清而言，則太極與太虛則有形而
　　　　　上下之分，此所以說：「非以太虛為太極也」。

二、王植對李注「太虛」義之批判

　　在〈臆說〉中，王植曰：「『太虛』二字是看《正蒙》入手關頭，於此得解，以下迎刃而解」。[22]因此對太虛三層義的分析，即是《正蒙初義》用以衡定所收各《注》的依據。他亦依此批判李注本為調和朱熹與張載之學，卻使「太虛」失去在《正蒙》中應有的本體義。以李注解「太虛無形，氣之本體」（〈太和〉）為例，李光地先批判前人之誤，在於以《太極圖說》的太極附會《正蒙》之太虛。另一方面，他則辨析太極與太虛之異，在於太極是形上之理，而太虛僅是無形之氣。二者存在著「形上／形下」的存有論區分。在《注解正蒙》中，「太虛」或可指向清通遼闊的「天空」或「虛空」，意指吾人經驗所見的天空雖然不具特定形質體貌（無形），甚至能以空曠涵容萬物的方式，包納一切氣化所成之物。[23]但

---

[22] 〈參兩〉，《正蒙初義》，卷2，頁2。
[23] 李注絕非第一個將「太虛」指向蒼蒼之「天」者，程顥（明道，1032-1085）即曰：「雖堯、舜之事，亦只是如太虛中一點浮雲過目」。王守仁（陽明，1472-1529）甚至表示：「良知之虛即是天之太虛」。至於陽明的良知本體與太虛的關係為何？以及是否由此帶出太虛的形上

無形之太虛，既非佛老所說的「空」或「無」，也非多數《正蒙》詮釋者所主張的形上理體。依此，李注對〈太和〉此句的詮釋可理解為：

在無形卻實有的太虛（虛空）中，已然存在著貞定氣化之所以然的本體，此即是形而上的太極。

但是王植卻反對李注的論點，提出兩項批判：

諸家以「太虛」、「太和」分體用，蓋以第一層與第二層對言也。無形處即是氣之本體，李注謂：「太虛無形之中，氣之本體存焉，即太極也」。「無形」、「本體」說作兩層，不免混入程朱甲裏，非張子本意。[24]

首先，王植指出，過往《正蒙》注解者多以「體／用」論架構，辨析太虛與太和之異同。但他認為張載由「太虛即氣」以明天道本體與造化之用，太虛即是湛然無形之氣，此氣雖不可即視為形而下之氣，卻本是貫通形而上下的氣化本體。故太虛本即兼含體用，而王植以太虛第一二層義串言之。由前論可知，王植亦以此為張載所見止於「太虛」，而論道體尚未明徹之處。但《正蒙》諸注家以「體／用」分言「太虛」與「太和」之義，卻將太虛本應串言之一二層義，割裂為相對的兩個概念。此詮釋上的關鍵滑轉，廣泛出現在包含李光地在內的《正蒙》諸注中[25]。其次，王植認為，李注將「太極／太虛」說成「氣之本體／無形之氣」或「形上／形下」的兩層關係，這樣對本體的詮釋是屬於伊川朱子式的「理／氣」論架構，而非張載思想脈絡中的虛氣關係。依此，王植曰：

意義？可參閱丁為祥與杜保瑞先生詳盡的探討。（宋）程顥、程頤，《二程集》，頁 61；（明）王守仁，《傳習錄下》，收入《王陽明全集》第一冊，卷 3，頁 117；丁為祥（1994：154）；杜保瑞（2012：40）。

[24] 〈太和〉，《正蒙初義》，卷 1，頁 12-13。

[25] 〈太和〉，《注解正蒙》，卷上，頁 1。

愚按此節首句即太虛第一層正義也。聚散變化，又言及第二層。「無形」即上「中涵之性」，「變化客形」即上所謂「浮沉、升降、動靜、絪縕、勝負、屈伸也」。「至靜無感」，即人心之太虛無形。「物交客感」，即人心之變化客形也。然有無、虛實通為一物者，性也。故曰：「盡性者一之」。蓋虛靜之中，事物之理無不具，必事物各盡其道，而後無形無感者，克全其體，此能一之義也。

又曰：

又張子言「性」與「神」處，俱以「性」屬第一層，「神」屬第二層。觀末篇「感者，性之神；性者，感之體」等語，可見《補注》皆以為氣之體，亦未確。太虛之義分三層足以貫之，而其立言不一，皆當隨文體認，細玩篇中之意。惟「由太虛有天之名，合虛與氣有性之名」，此若必以理言之，方可通。然亦張子所見止至太虛而止。若遂以為生物之理之盡乎此也者，而實未盡也。故朱子雖亦以「理」字釋之，而亦謂其生受辛苦，聖賢便不如此說。蓋非「理」字不足為言，而直以「理」字代之，則不可也。若「太虛不能無氣」、「萬物不能不散而為太虛」、「氣之聚散於太虛」、「太虛為清」、「萬象為太虛中所見之物」此類，如以「理」字訓之，則將謂「萬物散而為『理』」，「氣之聚散於『理』」，「萬象為『理』中所見之物」，於說通乎？又如「太虛無形」、「虛空即氣」、「虛能生氣」、「氣塊然太虛」、「太虛妙應之目」、「氣本之虛則湛」，此等直以「氣」言似未盡，而亦何可直以「理」字代之也？

他從兩面批判李注：

　　其一，以「太虛」一二層義串言，批判李注本欲辨析太虛與太極之
異，卻曲解張載以太虛詮解道體之義。王植認為，張載所言：「太虛無形，
氣之本體」，既是屬於太虛的第一層義（亦即「本體義」），同時又兼攝第
二層義（亦即「流行之用義」）。他反對李注取消太虛之本體義，反而在
太虛之上，另立太極作為萬物之所以存在的本體。李注的詮釋方式不僅
有疊床架屋之嫌，更違背張載「太虛即氣」的思想脈絡。他表示，太虛
之「無形」並非從現象上指明太虛本無形貌，而是在形而上的意義脈絡
中，顯明「太虛」並非為「形」所侷限之「物」，而是萬有之根源、造化
之本體。[26]在此又包含三項論述步驟：首先，他以推本窮源的方式指出，
凡有形之物皆歸本於無形，然若「無形」即是「虛無」，則張載之言將無
異於佛老崇虛尚無之論。故在張載思想中，無形之太虛必為拒斥「虛無」
之「實有」。再者，若將太虛視為超越於萬物之外而創化萬物的實有，則
顯然違背張載以虛氣本一的論述脈絡。為避免將「造化生生」拆解成「理
為生生之體，氣為生生之用」之弊。在他看來，太虛即是「生生」，並非
在氣化生生之外另有一作為氣化本源的理體或太極。依此，他從太虛本
是「兼含體用」而非「體用相對」，將太虛分解為第一與第二層義。並以
「生生」統攝太虛之體用義，而由此帶出太虛之第三層義：「原始反終」。
其意即是說，不論是從二氣絪縕看待整體生生歷程，或是就個體之死亡
而言客形之形潰反原，皆標示出道體並非孤絕於萬有生化之外的超越者
（第一層義），太虛即涵蘊「生生」（第二層義），且萬物出於太虛並復歸
太虛（第三層義）。故王植以「即體即用、原始反終」串言太虛之三層義。
最後，他從詮釋上的融貫性指出，將「太虛」視為形而上之本體，不僅
同樣能首尾一貫地解釋《正蒙》論及虛氣關係的文句，更能避免李注疊

26 　〈太和〉，《正蒙初義》，卷 1，頁 2-3。

床架屋之弊。故他批判李注對《正蒙》「太虛」義的解析是「混入程朱甲裏，非張子本意」。

　　其二，王植認為太虛與太極實各有能自圓其說的理論脈絡，應當將張載與程朱的思想分為各自獨立的兩種理論，而不應當勉強將二者混為一談。他承繼二程朱子對於張載的評斷，認為橫渠論究「道體」之處，總是苦心力索，卻「終是生受辛苦，聖賢便不如此說」。[27]因此，張載以太虛論說道體畢竟不如程朱以「理」直探形上本源來得簡明透徹。但若依此將程朱所言之「理」取代張載所言之「太虛」，則是扭曲《正蒙》文意。為澄清此觀點，王植採用概念替代的方式，指出前人妄自以「理」或「太極」詮釋「太虛」之誤。他以「太虛不能無氣」、「萬物不能不散而為太虛」、「氣之聚散於太虛」、「太虛為清」、「萬象為太虛中所見之物」等句舉例，若是將「太虛」詮釋為「理」可通，則以上這些文句當可代換為「『理』不能無氣」、「萬物不能不散而為『理』」、「氣之聚散於『理』」、「『理』為清」、「萬象為『理』中所見之物」。然而代換的結果，不僅造成文句的不通順，更在義理上難以說得通。依此可知，以「理」詮釋「太虛」並不恰當。同樣的方式，亦適用於反駁以「太極」取代「太虛」的詮釋。

　　基於上述，王植認為李注雖有見於前人以「太極」詮釋「太虛」之誤，卻反而將程朱理氣論之框架套用於詮釋《正蒙》虛氣論。如此不僅使太虛之本體義流失，也使張載虛氣論中理當澄清之處隱而未顯。[28]

---

[27] 《朱子語類‧張子之書二》，卷99，頁3328。

[28] 在注解〈太和〉所謂：「氣之聚散於太虛」一段時，王植曰：「諸家謂氣散歸虛，與下參伍變易之義不合，而理亦未瑩」，並指出：1.張載所言：「無無」之第二個「無」字，乃是指「太虛」，用以與下文「有無」之「無」字相對應，其目的在於「闢二氏有生於無等語也」。換言之，此處之「無無」，是藉由無形之太虛指明「虛無」並不存在，亦即從太虛之本體義，以「實有」拒斥「虛無」。2.但由《正蒙初義》所收錄之諸家注可見，注者多僅是以「氣散歸虛」詮解此句，未曾辨明氣之聚散於太虛，實是太虛之第二與第三層義，而非太虛之第一層

肆、王植對《注解正蒙》「神化」觀之論析

　　對於何謂「神化」？張載本自有定義：「神，天德，化，天道。德，其體，道，其用，一於氣而已」（〈神化〉）。當代某些學者認為，橫渠如此分解「神／化」二義，自有理論上的必要。[29]但亦有學者表示橫渠區分過細，反形成概念上的混淆（杜保瑞，2005：56-57）。王植對李注的批判，亦起於李注對「神化」之區分與新詮。

　　由前述可知，李注不僅洞見朱子以理詮釋太虛並未合乎《正蒙》之意，他更嘗試調和朱張之學，使二者論道體之異並非呈現儒學內的兩種對立理論，而是在同一儒學義理架構中，對道體的不同層次論述。因此他取消太虛的本體義，而將朱子所言之「理」另立在張載氣論架構的頂層，由此形成「形上之『理』」與「形下的『太虛』與『氣』」的存有論結構。李注認為，程朱對道體的洞見無可置疑，故形而上者僅能是「理」。但他也認為張載對造化（神化）之解析，更甚於伊川。因此他雖在「形而上／下」的架構中，將「太虛」視為形下之氣，卻認為太虛乃清通無礙之氣，為一切氣化所成之客形的本然樣態。此即就氣之「本然」，而言「太虛」即「氣」之「本體」。又由於一切現象皆於太虛中所呈現，而氣化所成之「物」（具體個物）與所呈之「跡」（漸、順、動、止等），皆以太虛為本體。故李注又由此帶出太虛即為覆地之天、蒼蒼者天等意涵，以太虛即為一切存有者與存在現象得以實存的存在場域。故太虛不僅是

義。依此，本段本應是聚焦在張載以太虛之本體義，反駁佛老形上思維中的「空」與「無」；但包含李注在內的注解者，卻僅注意到太虛作為流行之用與究極之歸的氣化活動。這顯是詮釋《正蒙》論述的滑轉與失真。

[29] 唐君毅先生指出，橫渠此分言「神」與「化」，各有兩串繫屬概念，本「互相對應，各有其義，不可混濫」。他甚至認為，朱子之所以肯定神化二字唯是橫渠推出來，正是基於此（2004：105-106）。

氣之本體，且貫通於有形與無形。現在的問題是，當李注調整了太虛在
《正蒙》中的存有論位階後，對於張載所言之「神化」顯然也必須做相
對應的調整，否則勢必形成理論上的衝突。此相應的調整即是李注神化
觀的特色，也成為王植批判的焦點。

一、李光地對《正蒙》「神化」義之詮釋

李注不僅接受橫渠以神化會通儒家經典的詮釋方式，而且嘗試由神
化觀調和伊川對橫渠的批判。[30]其要點有四：

其一，以「理／氣」關係詮釋「神／化」義。在論及《正蒙》「神化」
時，有兩處是理當辨析者：1.「神化」之「神」與「鬼神」之「神」是
否相同？2.「神化」是僅就「氣」上立論，還是就一氣流行中釐析出「理
／氣」兩層？關於前者，李注並未如同當代某些學者，將神化之神視為
「形而上」的太虛神體，用以區別僅屬「形而下」之氣化上事的鬼神之
神。[31]他認為：「神，惟一也」（〈神化〉），「鬼神」與「神化」之「神」，
皆為形上理體在氣化發用所呈顯的不同面向，故又以天之良能凸顯其流
行發用。[32]至於後者，李注既區分「神化」與「氣化」之不同，又將「神」
指向形而上的理體，[33]故「神／化」當是從氣化流行中可釐析為形而上

---

[30] 在注解〈太和〉之「太虛為清」段時，李注先提及：「程子議之曰：『神氣相極，周而無餘』，謂『清者為神，濁者何獨非神乎』」，並表示：「愚謂程子之言當矣」。但他接著指出：「然張子方言虛空之即氣，則有無、顯隱、神化之無二，斷無判清濁離神氣以自背其說之理」。〈太和〉，《注解正蒙》，卷上，頁8。

[31] 例如，牟宗三先生即表示：「（神化）同是表示『本體、宇宙論的』道德創造之體用不二、既超越又內在之充實圓盈之義。化之實、化之事，雖就氣說，然必于氣之虛實、動靜、聚散、有無、兼體而不累，參和而不偏，而見出神，始可成化。故《易傳》曰：『窮神知化』，簡言之曰：『神化』。說『氣化』乃只就化之實、化之事而言耳」（1990：冊1，474）。

[32] 〈太和〉，《注解正蒙》，卷上，頁10。

[33] 例如，對〈神化〉之「神，天德；化，天道」，李注曰：「神者，主宰，故曰天德。化者，功用，故曰天道」。李注認為，「神化」之「神」乃是主宰氣化流行之所以然的理體。

下的「理／氣」兩層義。[34]值得注意的是，合此兩點，李注認為《正蒙》神化觀與其「兩一」思想既當相互參看，也當有所辨明。由於〈太和〉曰：「兩不立則一不可見，一不可見則兩之用息」，且張載明確指出此處所言之「兩」（兩體），即是「氣」之「虛實也，動靜也，聚散也，清濁也」。這對於主張以「理」詮釋神化之神，又認為此「神」字與「一故神，兩故化」之「神」可互通的李注而言，詮釋的困難之處有二：一方面，他必須對〈參兩〉明白宣稱：「一物兩體，氣也」提出解釋[35]。另一方面，他則必須說明「理氣的不離不雜」與「神化的合兩見一」，如何在義理上可以相互通同？對此，李注特別澄清曰：「以『一』為『太極』，『兩』為『陰陽』也。然神化之解，朱子以為至精，只是一物周行乎陰陽、屈伸、往來、上下之間，所以謂：『兩在故不測』。天下之事，一不能化，惟兩而後能化。雖是『兩』，要之，亦推行乎此一爾」。李注認為張載雖說：「一物兩體，氣也」，但既非僅從「氣」上而言「一物」，將「兩體」視為「一氣可分化陰陽，陰陽統體是一氣」，也不是由此將結論導向「氣」即是萬有的根源。[36]李光地指出，此句當理解為「一」是指「太極」，「兩」是指「陰陽」，由於太極必是遍在於陰陽二氣往來屈伸的絪縕活動中，故所謂「氣也」並非以「太極」、「陰陽」皆只是「氣」，而是指理體是在陰陽往來的活動中，呈顯自身主宰氣化絪縕之所以能順而不妄的主宰義、妙運義與不測義。二者，李注又由「神」之不測義，帶出形上理體的遍在義與化生義。李光地指出，當張載以「兩在故不測」注解「神」之義時，即已表明「神化」之「神」，既是指形上理體妙運於陰陽動靜之中而不離

---

[34] 劉儓亦有類似的看法，而對〈神化〉篇首之「神，天德」句，注曰：「神，以理言」。（明）劉儓，《新刊正蒙解》，卷1，頁163。

[35] 〈參兩〉，《注解正蒙》，卷上，頁14-15。

[36] 例如，劉璣注曰：「太虛之中有一物而兩體者，氣而已矣」；張伯行亦認為橫渠所謂「一物兩體」，旨在言「氣」。故對此句即注曰：「氣雖有陰陽之分，其實止是一氣」。（明）劉璣，《正蒙會稿》，卷1，頁29；（清）張伯行，《正蒙注》，卷2，頁8。

於萬物，亦是指出理體雖遍在於客形世界中，卻並非任何一物。此所以李光地在解〈神化〉曰：「神無方，易無體，大且一而已爾」時，既表示此段：「一故神，然惟其無不在也，故大。大，故無方」，指出「神」即是唯一且遍在的理體，由此凸顯神之本體義；後文又藉由其注曰：「兩故化，然惟其推行乎一也，故一。一，故無體」，再次強調形上理體正是以其無所不在，故能保障萬有之存在，正是以無特定之方體，故能於氣化推行中展現生生不息的妙用。

其二，闡明張載藉「窮神知化」發揮「性與天道合一」之旨。李注嘗試調和朱、張之異。例如，在注〈中正〉之「儒者窮理」段時，他不僅明確宣稱：「性即理也」，更表示：「不以理為性，則將以何為道而推而行之乎？」（頁73）。依此，當他從《中庸》「天命之謂性」詮釋〈太和〉所謂「合虛與氣有性之名」時，既嘗試從橫渠氣論中釐清「性」與「太虛」之本「一」關係，又進而提出了朱子學的說明：「此天之所以命人之性也」（頁9），將「性與天道」之合「一」，視為「義理之性」墮入「氣質」之中的關係。基於人性可推本於天道，而天道又是天地萬象之形上理序。故在李注的天人架構中，人可據此義理之性，全然無隱的領會天地整體之理序。[37]他雖亦言：「形而上者，不離乎象，而不可象也」（頁32），以為天道之全幅義蘊非吾人「見聞之知」所能盡知[38]。但李注承接張載對「天德良知／見聞之知」的區分，並由性即理詮釋橫渠所謂：「性，

---

[37] 〈神化〉，《注解正蒙》，卷上，頁37。

[38] 李注點明《正蒙》將人之「知」區分為「德性之知」與「見聞之知」的理論意義。他指出，吾人若非無從全然得知天道理序，則是以一種有別於見聞的方式，體證此形而上之理。對〈大心〉之「德性所知，不萌於見聞」，李注曰：「萬物皆備於我矣！何則？其性與我同出於天也」（頁57-58）。值得注意的是，他不僅認為天理非得自於人對外在世界的觀察，甚至主張：「物性之神，亦非自外得也」（頁25）。由此將「格物致知」從對物理自然的客觀觀察，一轉而為哲學的觀物智慧。

天德；命，天理」（〈誠明〉）、「性天經然後仁義行」（〈至當〉），將吾人稟受於天的義理之性，視為人之所以能領會天道理序的本然根據，人之德行乃源自於對此性的真知與奉行。所謂「窮神知化」，並非意指人當從自然世界的經驗觀察中，歸納出各種運動變化現象的物理歸準或實然法則。因為此僅是知曉天道之跡，而非天道之所以運行之理。「窮神知化」不以外求為本，而是以人之性體為實。此所以在注〈王禘〉之「至誠，天性也」，他再次表示：「盡性，則可以窮神；至命，則可以知化」，由此呼應〈誠明〉所謂：「德不勝氣，性命於氣；德勝其氣，性命於德」，[39]將人依義理之性而超越氣質之限的實踐工夫，視為「成性」工夫。

其三，由「過而不留不滯於物」詮釋「存神過化」。對張載所謂：「存神過化，忘物累，而順性命」，李注曰：「張子言過化，猶言過而不留不滯於物也。故忘物累，則不徇於物，不徇於物，則不化於物。過化之道也。順性命，則不喪其心，不喪其心，則不滅於理。存神之方也」。在注文中，他先呈現張載在此實吸收了莊子「物物而不物於物」的表述方式，進而轉化《孟子》論「過化存神」之文脈義旨。《孟子》本以此表示君子之德行蘊含無邊的渲染力，能使周遭，甚至四方之人，皆能受其感化，自願的追隨君子而行。〈滕文公上〉所言，實即闡發〈顏淵〉「風行草偃」之旨。[40]張載則從形而上的進路開展孟子學。先以氣論展開「天人合一」的架構。次由「神化」說明「天道四時行，百物生」（〈天道〉）之至德與創化活動。再以萬象皆神化之糟沫，指出天地萬物一體相連的本然關係。最後，則依此說明君子何以能跨越人物氣質之隔，以德行感通人我，收存神過化之功。李注則點出《正蒙》善藉《莊子》之表述以詮釋儒家義

---

[39] 〈誠明〉，《注解正蒙》，卷上，頁 52-53。
[40] 《四書章句集注》，《朱子全書》，第 6 冊，頁 306-308。

理的特色，從「物物而不物於物」[41]闡發儒學德行工夫如何超拔於「徇物喪心」，進而在整體德行實踐中，保持天德良知的靈明，遵行義理之性的指引。由此緊扣〈太和〉「神化性命通一而無二」的義理架構。換言之，李注再次從「性即理」指出天道與人性的一本關係，並且將《易傳》之「窮神知化」與《孟子》之「存神過化」，皆視為古代聖賢從不同面向闡釋此天人合一之理。依此，注〈神化〉「性性為能存神，物物為能過化」時，他徵引《周易・繫辭上傳》之「成性存存，道義之門」呼應此處所揭示的孟學義理，而曰：「性性者，成性存存也；物物者，以物付物也」。

　　其四，綜論《正蒙》神化觀如何貫穿儒家經典要旨。儘管「經」在中國文化裡擁有超越於史、子、集類的正典地位，但當理學家徵引儒家經典以闡發義理時，往往並未嚴格遵守這些文獻本來的行文脈絡，而是採取一種「斷章取義」的方式，呈現其對儒家各經典之共通義理的洞察。在注〈神化〉之「氣有陰陽，推行有漸為化，合一不測為神」段時，李注即藉此指出張載用以貫穿不同儒典的這種洞察。李注認為，「神化」可由「在天」與「在人」分別立說。在天者，即橫渠由氣化發用處指出，使氣化之所以然之「神（理）」，不離且不雜於氣化流行，由此發明天道生物之神妙不測。在人者，即就德盛仁熟者之體段與工夫處，闡明致學成聖者得以體察天道生物之理的本然依據，以及實踐進路。由於道體本獨一無二，故古今聖賢雖在體證道體時有不同的入路，而在撰寫經典時呈現不同的論述。但經典中所傳達的義理，理當是通一無二的。他舉例表示，《易傳》之「精義入神」、「窮神知化」，是從人體證天道生生之德處而言。故其要旨有二：一者，闡明人之性體即為人之所以能體察天道

---

[41] 例如，《莊子・山木》曰：「物物而不物於物」（頁 668），〈知北遊〉曰：「物物者與物無際」（頁 752）。對〈神化〉所謂：「性性為能存神，物物為能過化」，李光地更直接以「性性者，成性存存也；物物者，以物付物也」詮解之。（清）郭慶藩，《莊子集釋》；〈太和〉，《注解正蒙》，卷上，頁 38。

之神的依據。二者，揭舉人如何體道證德的實踐功夫。前者即《中庸》首章所謂「天命之謂性」，後者則呼應《論語・憲問》「下學而上達」之道。又如同孟子稱揚孔子乃「聖之時者」，自是由於孔子之德行能變通不滯、與時偕行。但孔子之所以能知「時」而「與時偕行」，李注則嘗試以神化會通孔孟義理，認為這是推本至聖人體察天道神化之理處而論。[42]

二、《正蒙初義》的神化觀

王植以太虛之三層義詮釋「神化」，有兩項重要的主張：其一，就「神化」與「太虛」之關係，二者並非兩個本質相異的概念。「太虛」兼含體用義，「神化」是指向太虛本具流行發用的活動義。〈臆說〉即曰：「其言神、言化、言命，皆即太虛之流行者，而分別言之，蓋寂然不動則一而已，感而遂通乃有三者之異名」。[43]就「神」與「化」合而言之，則《正蒙》凡言及神化之處，皆是從太虛之第二層義上論說。「神化」皆指向天道生物以後之事，涵蘊造化生生之妙用。其二，就「神／化」析解區分而言，「神」屬於「氣之用之體」，而「化」屬於「氣之用之用」。王植綜觀明清《正蒙》諸注而指出，在理氣關係中，「理」為「體」，「氣」為「用」，屬諸儒共識。[44]但張載有別於諸儒者，正在於他以「氣」亦自有「氣之體」與「氣之用」，而且「氣之體用又各自有體用之可分」。依照王植的解析，《正蒙》氣論中的體用關係可分解如下：

---

[42] 〈神化〉，《注解正蒙》，卷上，頁 33-35。

[43] 〈臆說〉，《正蒙初義》，卷 1，頁 3。

[44] 誠如楊儒賓先生所指出，自朱子詮解《正蒙》以來，以「理為體，氣為用」的方式詮釋張載氣論，已經是許多服膺程朱理學之儒者的共通見解。朱子的理氣論雖是明代中晚期反朱子學者所欲對治的目標，但對治與被對治者卻共同形成了包含正反兩面論證在內的論域，使得反對者也是繞著理氣關係展開論辯。鄭宗義先生則指出，朱子的太極之理是只存有而不活動，動靜的是氣。但朱子解〈太極圖說〉與《通書》時，「往往把太極之理、動靜、體用等合言，遂造成後學一些不必要的滑轉與誤解」。這樣的現象，在後儒援引朱子之注解探討《正蒙》哲學時，亦不罕見（楊儒賓，2007：254；鄭宗義，2009：117-118）。

1. 「氣之體」，就「未有天地人物之先」而言，「氣」初僅指向存有為真實無妄之「有」而「非無」，萬物皆尚未由太虛中分化，此為「氣之體」。王植以「無形者，皆體也」，標示「氣之體」為氣之本然，「無形」並非無有形體，而是尚未為「形」所分化，且不為「形」所侷限。再者，王植以為「氣之體」又可依氣之本然的本體義與發用義，再區分為「體之體」與「體之用」。李注正是脫落此義。

   （1）「氣」之「體之體」。凡言及天地萬有之本然，進而直指存在根源與價值理序者，屬「氣之體」中的「體之體」，此即「太虛」之本體義。例如，太虛即是氣之「渾然無朕之本體」，故為「氣之體」之「體」，簡言「體之體」。

   （2）「氣」之「體之用」。就氣之本然雖無形無象卻非靜止之死理而言，未曾分化之氣已然涵蘊造化生物之大用，故可分屬為「氣之體」中的「體之用」，此即「太虛」之發用義。氣由「升降飛揚」、「相摩相盪」顯生物之「大用」，此即「氣之體」之「用」，簡言「體之用」。

2. 「氣之用」，就「既有天地人物之後」而言，則一氣已然分化為萬有，王植以「有形者，皆用也」表示形器皆是氣化之發用流行，故稱之以「氣之用」。再者，由於個別存有者是在整體氣化歷程中以游氣聚散的方式生滅，故相對於整體存有歷程而言，個體生成亦屬「氣之用」。三者，王植以「氣之用」亦可再區分為「用之體」與「用之用」。

   （1）「氣」之「用之體」。王植將一氣已然化生萬有後的氣化活動再做分解，若直指此氣化活動本即是以清濁聚散、沉降飛揚的方式發用流行，則為「氣之用」之「體」，簡言「用

之體」。由於此時所指之氣已非氣之本然，而是一氣分化以
後之事。故雖直指此活動之本然而名之為「體」，卻是已然
有分、相互對待之「體」，故王植又以「用之體」為「對待
之本體」。

（2）「氣」之「用之用」，則是指風、雨、雪、霜等天地萬物與
萬象，皆不過是此氣化流行之各環節又經由分化、交感所
化生，故王植又稱之為「流行之妙用」（頁27）。

依此體用觀詮釋「神化」，在注〈太和〉之「神化、性命通一無二」
段時，王植曰：「神以太虛流行之妙言，即首節清通之神。性與命對，神
與化對。命是性之流行，化是神之發用。以性與神對，則性屬第一層，
神與命與化皆在第二層內。」（頁22）。他不僅將「神化」統體視為太虛
第二層義，屬「氣之用」。並且就「神」與「化」之區分而言，「神」是
「用之體」，而「化」屬「用之用」。[45]王植善於將概念層層釐析，在探

---

[45] 由上述可知，王植認為，朱子以「理」為道體，「理／氣」為「形而上／下」且「不離不雜」
的關係，才是確當之論。但他亦指出，《正蒙》諸注解者以為張載氣論亦是如此，則有誤將
橫渠理解為程朱學脈之嫌。故他先以上述架構分析張子之太虛，以為《正蒙》之精采與侷限
皆在氣論上。但順之而來的問題即是：設若「有象者，皆（氣之）用也。然則理於何附？」
對此，王植答曰：「順而不妄、不得已而然者，即『理』之所以即氣而存，而不雜乎氣者也」
（頁27）。他甚至認為，此順而不妄的天地之理，不僅是人之所以存在的形上根源，也是天
所賦予「我」，而使「我」地能全然領會天之理序、與天合而為一的內在依據，此即吾人本
然稟受於天之性。此所以王植曰：「塞吾其體，帥吾其性，故得天地之氣為周身之氣而形生
焉，得天地之理為浩然之氣，而性成焉。此人之所以可達於天也」。他在此不僅強調〈太和〉
氣論與〈西銘〉仁孝之理不可相互分割的理論一體性。更將張載在〈誠明〉中的成性觀與朱
子「性即理也」之論相互會通，嘗試調和二者在人性論上的差異。值得注意的是，王植既然
以程朱為「正學」，又為何不用心發揚程朱之學？還專注於批判後儒將《正蒙》「程朱化」呢？
本文以為，王植的用心有二：其一，張載論道體雖不似程朱透徹，但《正蒙》苦心極力以闡
發道體之論仍有獨到之處。例如，「一神兩化」之論，正是張子論道體精采之處。依此，吾
人可批判張載論道不足之處，但將《正蒙》程朱化，則是連張載獨到的見解一併抹殺。對此，
可由《正蒙初義‧序論》徵引范育之〈序〉，以答二程朱子對張載的批判中得見。其二，《正
蒙初義》不僅是對多種《正蒙》注釋的靜態集結，而且是從《正蒙》注解的動態發展中，建
立起一條澄明儒家義理的論道進路。換言之，王植是以《正蒙》注解史撰寫理學發展史。依
此，張載見道體之未足，反而更能突顯程朱何以在理學史上應為「正學」。

討「神化」概念時，尤其如此。他不僅辨析「太虛」、「神化」，以及「性」、「神」、「化」、「命」等概念的同異關係。他甚至認為，吾人「須於無層次中分出層次看」，[46]方能掌握《正蒙》義蘊。再者，王植不僅從體用關係區分神化之不同義蘊，他還依據文脈，將「神化」釐析為「天之神化」與「人之神化」。所謂「天之神化」，乃是從天道之化運而言，此即〈神化〉曰：「氣有陰陽，推行有漸為化，合一不測為神」。所謂「人之神化」，乃是指人應時時無忘在德行修養實踐中，追求達至德熟仁盛之境，由此能領會奉行天道於氣化之象中所呈顯的形上理序，此即〈神化〉所言：「其在人也，（知）〔智〕義（用）利〔用〕，則神化之事備矣」。[47]依此，王植一方面曰：「可見天有化，人亦有化。天之化也，運諸氣，以為流行。人之化也，順夫時，以為變化」。另一方面，他更表示：「天非氣，人非時，則『化之名』於何有？『化之實』於何施手？」，[48]闡明橫渠如何以「神化」展開「天人合一」的氣論論述。值得注意的是，王植區分「天之神化／人之神化」的另一項關鍵意義，在於藉此釐清張載徵引儒家經典時，是考據《中庸》等經典之本義？或是藉由《論》、《孟》、《易》、《禮》之言，發揮自身對古代聖賢共通義理之理解與詮釋？若屬於後者，張載所闡發之義理又當如何與其氣論相互串言？基於此，王植在此處的區分，既成為他對橫渠經典詮釋原則的理解，也成為他批判《正蒙》注解者是否掌握張子引儒家諸經的標準。例如，他指出：「《中庸》曰：『至誠為能化』，《孟子》曰：『大而化之』，皆以其德合陰陽，與天地同流而無不通也」，他依此批判包含李光地在內的注解者，在注釋〈神化〉廣引諸經時，陷入頗多牽合、反難安頓的困境中。[49]

---

[46] 〈神化〉，《正蒙初義》，卷 4，頁 3。
[47] 〈神化〉，《正蒙》，收入《張載集》，頁 16。
[48] 〈神化〉，《正蒙初義》，卷 4，頁 10。
[49] 〈神化〉，《正蒙初義》，卷 4，頁 11-12。

三、王植對《注解正蒙》神化觀之批判

由前述可知，李注認為，張子神化旨在推天道以明人事，闡發天人本無二之理。他表示：「天下之事，一不能化，惟兩而後能化。雖是『兩』，要之，亦推行乎此『一』爾」。他首先從「理／事」二分的架構銜接「理／氣」論，以「天下之事」將原應探究本體的論述轉向作用層，並依此聚焦在事物之理上。其次，他則由「能化／不能化」指出，除非吾人所見之世界乃一死寂靜默的世界，否則天地萬象之活動即先行的預設了作用變化之所以可能的原理。此就《正蒙》而言，即是由兩端對反中見統一活動整體的「兩一」觀。基於此，李注對「神化」提出了略顯繚繞的說明。此即王植批評的關鍵。

（一）、王植對李注「一神兩化」說之批判

在注解「一神兩化」時，王植便指出：「一神兩化以首篇之義推之，俱在太虛第二層內。蓋兩而化者，固二氣之運；一而神，亦止就發用處言也。『太極兩儀』之云，濂關所論有不可強同者」。他反對學者強將太極與太虛等同，亦順此反對以「理／氣」二分詮釋「神化」的觀點。在他看來，「神化」只是就太虛在作用層上的意義立說。若勉強會通濂溪與張載之學，強分「神／化」為「太極／氣化」，則既是混淆太虛之本體義與作用義的區分，也使太虛與氣本屬一體的關係，硬被割裂為「形上／形下」不可逾越的關係。因此對於《正蒙補注》在內的這種詮釋觀點，他便批判是：「已入《太極圖說》窠裏，非張子本意」。[50]再者，李注雖宣稱自己是承繼朱子的詮釋，將「一，故神；兩，故化」解析為「理（神）不離且不雜於陰陽二氣之妙運（化）」。然而，李注與朱子並不相同。朱

子雖以張載未能全然洞悉道體之實，仍以「太虛」旨在言「道」而非「氣」。但李注以「太虛」雖為清通無礙之氣，卻只言及「氣」而非「道」。設若「神化」之「神」，是指形而上之「理」，而「化」指形而下的陰陽二氣之運行。在朱子的詮釋中，「神／化」乃可指向「太虛（理）／陰陽（氣化）」。[51] 此詮釋雖不同於張子原以「一物兩體」乃是「太虛之氣」兼含「陰陽、動靜、屈伸二體（體狀）」，太虛之本體義仍未脫落。但在李注中，「神」與「化」分指「理」、「氣」，且「太虛」非「理」，故「太虛」不指向「神」。順此脈絡，除非李注以「太虛」為「神化」之「化」，否則「太虛」在「神／化」之架構中，將無所對應。但依李注，「太虛」僅為「清通無礙之氣」，屬形而下層，既不足以概括陰陽二氣，也非氣化之妙運。故李注不僅不合乎張載之原旨，也並非朱子所言之意。故當李注在詮解「神化」與「太虛」之關係時，得援引〈太和〉之「氣之聚散於太虛」，將「太虛」視為含納陰陽二氣之聚散的場域。

（二）、王植批判李注未能辨析橫渠神化觀之經典詮釋原則

　　王植批評李注藉由「神化」會通《中庸》之「敦化」與《孟子》之「過化」。由於在解釋一神兩化時，李注已將神化導向人事之理上。依此，李注認為，當張載援引儒家經典與神化觀相互發明時，有三點值得注意之處：一者，張載並非以考據訓詁《論》、《孟》、《中庸》等經典為首要關懷，而是藉由梳理不同經典中的共通義理，闡明儒家天道性命之學。因此，橫渠門人雖曾集結《經學理窟》，但張載經學往往是隨文發揮，而

---

[51] 在朱子思想中，「神」可指向「氣」，例如，「鬼神」之「神」。但朱子並不認為張載所言之「神化」即等同於「氣化」。從他辨析「太虛即氣」可知，他所批評的是張載本欲言形而上之理（或太極），卻總在論述中滑降下了氣邊之事。依此，他嘗試藉由嚴分「形上／形下」的理氣論架構，替張載澄清論說未明的神化、虛氣關係。《朱子語類・張子書一》，卷 98，頁 3302-3303；〈張子書二〉，卷 99，頁 3335-3336。

超出經典本有的文義。[52]在闡發「神化」觀時亦同。其次，李注認為張載以神化詮釋《中庸》之敦化，是從「體／用」觀上，以「敦」標示出「人」本是受命於天的道德主體，而由「化」指出人在道德實踐上應有的判斷力、應變力與感化力。合此二者，「大德敦化」即意味聖人之德是能領會氣化流行中所呈顯的當然之理，並在人倫日用中廣泛應用而造福人群者。因此，「神化」即由理氣論論述，通過銜接「敦化」之詮釋，轉向屬於仁智之事的德性實踐工夫。再者，李注認為張載以「神化」解釋《孟子》之「存神過化」，亦是指向人涵養本受於天的道德性體（「所存者神」），故能在人倫日用之中，不受感官欲望與外在事物的引誘牽絆，得以時時依循天理而行，此即以「過而不留，不滯於物」詮釋「所過者化」。[53]依此三點，李注認為，《中庸》之「大德敦化」與《孟子》之「存神過化」，正足以闡明儒家天道性命之學，皆指向有德者依據天所與我的道德本性，由下學上達、窮神知化，既領會天道之神所呈顯的天理秩序，又能「變通不滯，與時偕行」。然而王植並不認同李注的詮釋，而從三方面提出辨析：

其一，他援引真德秀（西山，1178-1235）對橫渠的批判，指出李注承繼《正蒙》之「性性為能『存神』，物物為能『過化』」，實是對《孟子》的過度詮釋。[54]《孟子》言「過化存神」原是對聖人以德化民而民自化

---

[52] 《經學理窟・義理》即記載張載的經典讀書法，曰：「觀書必總其言而求作者之意」，「有言經義須人人說得別，此不然。天下義理只容有一箇是，無兩箇是」（頁275）。至於張載對經典之詮釋與理解，以及在此中所呈現的詮釋學式反省。可參閱林維杰先生深入的探討（2012：35-40）。

[53] 〈神化〉，《注解正蒙》，卷上，頁37-38。

[54] 真德秀（西山，1178-1235）曰：「《孟子》『過化』謂聖人凡所經歷處，人皆化之；『存神』謂其中所存神妙，正意只是如此。至橫渠先生，乃謂『性性為能存神；物物為能過化』。下性字指本然者而言；上性字謂我能全其性，而不為情所蕩，則其所存神妙不可測。下物字指事物而言；上物字指我之應物而言，謂物物各自有理，我隨其理以應之，物各付物不以己之私意參乎其間，則事過弗留，如氷之釋，如風之休。後來諸老先生多本此說，獨文公不以為然者。蓋孟子之意未說到如此深也。文公解經平實如此。然橫渠先生之說亦不可不知也」。

的描述，是以神妙不測強調德化活動之普遍性與必然性。但張載卻將《孟子》文意轉向天道性命之學，一方面從「性性」言人之道德實踐本有根源於天地的形上根源。另一方面，又轉化莊子「乘物遊心」之學，從「物物而不物於物」言人若依循天地之性而行，則能使心不迷失於物欲，達到駕馭氣質而超越氣質所限的境界。若順王植的批判可知，李注之誤並不在於指陳橫渠以天道性命之學發揮《孟子》、《中庸》之旨，而是李注更進一步將橫渠天道性命之學，導入伊川朱子「性即理」的脈絡中。故李注以「神」指向「天理」，「存神」意指人存養同於天理之性，而「過化」即是擴充此性以保持德性心之靈動，在洞見事事物物中所涵之天理。由此又將朱子「格物窮理」與「神化」觀銜接。至此，當〈大心〉以「大其心則能體天下之物」回應《孟子》「盡心知性以知天」之學時，橫渠所言之心，在李注的詮釋下，則由貫通天人的形而上之道德心，一轉而為在「能知／所知」對列格局下，能知（窮究）天理的認知心。

由引文可知有三：其一，真德秀承繼朱子的論點，認為張載詮釋《孟子》「過化存神」已經逾越典籍本有的論域範圍。《孟子》的意旨，僅是闡明聖人是以其無邊德行感化眾人，故聖人所經歷處，眾人皆受此德行薰陶而自願轉化自身行為，朝向更為美善的理想發展。「存神」是指此德化活動所呈顯涵蘊的功效實神妙難言，且其中又無絲毫人為造作。其二，張載藉「過化存神」發明己意，而將《孟子》自行詮釋為：人與生俱來的本然之性（天地之性），即是人稟受於天之內在道德根源，人若成全此道德性而不為氣質之性所侷限引誘，則人的生命歷程亦即是超越種種侷限以達至天人為一的道德實踐活動。故以「性性」詮釋「存神」，描述人之德性實踐本有與天為一的超越根源。人之實踐（現）善並非出於外力的規範強迫，而是承繼天之所與我，我固有之的善性，而成就此人之所以為人之性。至於以「物物」詮釋「過化」，則是發揮莊學之無心應化，由此將孟子之意，詮釋為：有德者能依此心所領會之天理與物物各自之理而應物，使此心不因己私而牽繫於物，在物各付物中，達到「事過弗留」、乘物而化之的功效與境界。其三，真德秀認為，張載雖過度詮釋《孟子》文獻，而且廣泛影響其後儒者解經的觀點。但由於張載之言實有獨到見地，故在解經時，雖必當辨明《孟子》與張載所言「過化存神」實有不同，但也應保留張載的詮釋，提供學者思辨經典的另種視野。然而誠如匿名審查人所指出：「西山所尊者程朱，但程朱解孟亦不是孟子原意」；孟子論「君子所過者化，可以指所經歷過之處，民風都能成其變化」。理學家在解讀《孟子》時，不僅常隨著體證天道性命的入路差異，而有截然不同的詮釋，更不乏依據體驗哲學式的隨文發揮，在詮釋《孟子》時溢出文本原有的文脈與論域。但礙於篇幅所限，本文此處無法詳細辨析張載、西山、李光地，以及王植等人，何者較貼近《孟子》之原意。〈神化〉，《正蒙初義》，卷4，頁28。

　　其二，王植認為，《中庸》之「敦化」與《孟子》之「過化」本有不同的論述脈絡，應當理清各自的義理而非強求會通為一。依此，王植不僅表示：「此節大意，蓋以《中庸》『敦化』之意合於《孟子》『存神過化』之文，而皆非其本意」，更指出：「『敦化』之『化』，作『變化』之意；『過化』之『化』，則『過而不留』之意，上下似難合一」。基於此，他不僅批判李注強同二者實有斷章取義之嫌，更接受《正蒙補注》的論點，認為「性性」、「物物」雖觸及「大德敦化」的文意，卻尚不足以窮盡「仁智合一」的境界。[55]李注既未能辨明橫渠與《孟子》文意之不同，又多作繚繞而未能盡聖人之意。故批評曰：「李注於難解之處，割裂最多。雖詮注易明，恐失橫渠本意，不若《大全》原本為善」。[56]

　　其三，王植批判李注以「神化」釋「所過者化」為「化物累者也」，有陷入釋氏「銷礙入空」之嫌。[57]王植認為，〈神化〉雖言：「無我然後得正己之盡」，言：「存神過化，忘物累而順性命」，但吾人理當綜觀全篇文脈而理解張子之意。他指出，張子本段「首三句言天道之神化」，至於「『其在人也』以下六句，言人之窮神知化也」，其前後文關鍵在於推天道以明人事，並且闡明「《易》言『窮神知化』，乃德盛仁熟而致之」（頁16）。依此，李注之非有二：一者，李注未辨明由「銷礙入空」與「捨惡趨善」而言「化」，在義理進路上實有儒釋兩家本質差異。王植曰：「銷礙入空，猶言由『有』入『無』，即『冰銷雪化』之『化』。言之出於釋

---

[55] 〈神化〉，《正蒙初義》，卷4，頁28。

[56] 〈臆說〉，《正蒙初義》，頁6。

[57] 誠如匿名審查人所指出，李注雖側重「過化」在德行實踐中的消融義與工夫義，但李注並未順此「解消工夫」有意銷除了超越的理體，或吾人之性體。設若「銷礙入空」乃是指：工夫論上的消解義實源於否定本體的真實性，而落入釋氏視天地萬物皆無自性之說。顯然李注並無此意。但王植見李光地以「所過者化，化物累者也」注「範圍天地之化而不過」，又見他釋「旁行不流」為「所過者化，則旁行而不流。徇物喪志，溺於流者也」，便批判李注有「銷礙入空」之嫌。本文亦以為，王植之批判反顯得有些矯枉過正。〈神化〉，《注解正蒙》，卷上，頁38-39。在此再次感謝審查人細心的提醒。

氏者也。舍惡趨善，即『變化氣質』之『化』。學者所恒言也」。二者，李注僅著眼於「化物累」，而窄化張載由「神化」串言「窮神知化」與「過化存神」，旨在闡明儒者參贊天地之化育，實以萬民之福祉為共通本懷。王植認為，《正蒙》從肯認天道之真實無妄上，闡明世間萬象不僅真實非虛，且天道即氣化流行所呈顯的元亨利貞之德，正是萬象生成流逝所奉行的價值理序。由此不僅從存有論的高度，駁斥老氏無中生有之言與釋氏以萬象皆空之論。更將孔孟所倡言的人倫理序，推本至天道秩序。人之德行實踐即對天道秩序的領會，與天命召喚之回應。道德生命之實現絕不僅止於一己是否徇物喪心，更在於如何通過我之有所作為，而促成天地萬物之美好和諧。因此對「人之化也，順夫時」句，王植注以「德合陰陽、天地同流，則順時之實也」（頁 11）。張子所言之「無我」、「忘物累」，即便在表述上接近莊子之言，但在義理上則純是儒家的進路。故他特指出：「『天之化』二句，又為下半節之綱」。注《正蒙》者，僅見張載以「忘物累」對治前句「徇物喪心」，便將張子所言在人之神化等同於解消物累之學。更有甚者，將釋氏以萬法空寂對治人生執著窒礙之言，附會為張載此處所論，遂以為銷礙入空即是〈神化〉以氣論串言存神過化、窮神知化的全幅義蘊。故儘管李注亦順張載之言而注曰：「銷礙去惡為化，此始學之化，非天地聖人之化也」，王植卻由李注後文仍以「過化之道」、「存神之方」皆旨在闡發人如何避免「徇物喪心」，判定李注並未掌握《正蒙》要旨。故批判曰：「學者之舍惡趨善以為化，此只可為始學。遣去物累者，淺乎言『化』之義云爾！豈可與天道之神化同日而語哉」。[58]

基於上述，王植批判李注詮釋「神化」觀之失。但他不僅指出李光地詮釋上的缺點，在論析「『神化』是否可分先後？」以及「『神化』是否承漸不承速？」的議題時，王植便認為李注的詮釋獨得張載之意，解

---

[58] 〈神化〉，《正蒙初義》，卷 4，頁 11。

出「雷霆」之喻旨在強調窮神知化的為德工夫不離厚積薄發的實踐進路，一如「雷霆」這項造化所呈現的自然現象，其發生於吾人眼前雖至為迅速，但實則亦是逐漸累積各項發生條件而成。[59]王植亦總結《正蒙》諸注的論點，而肯定的表示：「惟李注得之」。[60]

伍、結論

　　設若研讀哲學家的代表著作，是吾人探究其思想的必要條件之一，則逐句解析張載思想論述的明清《正蒙》注，顯然是研究橫渠思想在明清哲學之發展時，所無可忽視的一環。尤其歷來注解《正蒙》者，不乏高攀龍，王夫之等望重士林，或在哲學史上具有獨創見解者。然而在當代張載學研究中，相關議題似乎仍較少引起學者間的討論興趣。本文以王植對《注解正蒙》神化觀之批判為題，除基於「神化」本即是張載哲學中極為關鍵，且廣受學者討論的概念外，還包含兩項考量：一者，由於李注在清代張載學研究中的代表性；二者，則是源於《正蒙初義》不僅是收集明清《正蒙》注最詳的作品，王植對李注之徵引與批判也是可見資料中最為全面且深入者。又由上述討論，本文獲致以下三項結論：

　　其一，王植批判李注使「太虛」失去在《正蒙》中應有的本體義。這又包含兩方面：一者，王植先肯定李注以濂溪之太極與橫渠之太虛本不可強同，實發前賢所未見。但李注與多數《正蒙》詮釋者相同，當以「體／用」二分架構辨析「太虛」之義時，使本兼含體用二義的太虛概念，陷入體用二判的困境。二者，李注依前述將「太虛」僅視為流行發用之氣，而非形上本體。李光地服膺朱子對道體之發明，又肯定張載論

---

[59]　〈參兩〉，《正蒙初義》，卷2，頁39-40。
[60]　〈參兩〉，《正蒙初義》，卷2，頁40-41。

述造化之深切。故嘗試調和朱張二子之說，而將朱子所言之「理」另立在「太虛」之上。由此將《正蒙》詮解為「形上之『理』／形下之『太虛』（清通不可象之氣）與『氣』（散殊可象之氣）」的理氣論格局。但王植指出，張載雖論述道體未徹，卻仍以太虛發明道體之義蘊。誠如「太極」與「太虛」不可強同，以「理」取代「太虛」，亦無法順通《正蒙》論及太虛的所有文句。

其二，王植批判李注以「理／氣」二判強分「神／化」為二。由於「太虛」在《正蒙》中的核心地位，李注既取消太虛之本體義，勢必得連帶調整他對「神化」的詮釋。在李注中，「理」是形上本體，「太虛」雖清通不可象，卻仍是形而下之氣。故當他以「神／化」分指「理／氣」時，「太虛」不僅失去本體義，而且隨著「神」之義由太虛中剝離，「神」亦失去創化之能動義。

其三，王植批判李注未能辨析橫渠神化觀之經典詮釋原則。此又包含三項批判：一者，橫渠詮釋「存神過化」已然踰越《孟子》文脈所言，但李注未能辨明。至於橫渠以「神化」串言《中庸》之「敦化」與《孟子》之「過化」，但兩本經典本有不同的論述脈絡，李注亦未能辨析張載如此會通儒家經典之得失。二者，李注引朱子「性即理」架構詮釋《正蒙》天道性命論，不僅使朱、張論「義理之性／氣質之性」的本異反成渾一，更將張載心性論所言之形而上的道德本心，滑轉為朱子體系中的認知心。故在解析《正蒙》串言《易》之「窮神知化」與《孟子》之「存神過化」時，李注以朱子「格物窮理」之「窮」釋之，而有違橫渠實以「大其心」闡明道德本心之形上義與天心之無外義通同為一，而此「大」、「無外」之義方是「窮神知化」之「窮」與「知」之義。三者，王植批判李注以「神化」釋「所過者化」為「化物累者也」，不僅陷入釋氏「銷礙入空」之說，而且有窄化《正蒙》興滅繼絕的儒家本懷。

參考文獻

一、古籍

王夫之（明）（2011）。《張子正蒙注》。收入《船山全書》第 12 冊。長沙：嶽麓書社。

王植（清）（1983）。〈臆說〉，《正蒙初義》。台北：臺灣商務印書館。

王陽明（明）（2010）。《傳習錄下》。收入《王陽明全集》第一冊。杭州：浙江古籍出版社。

朱熹（宋）（2002）。《四書章句集注》。收入朱傑人等主編，《朱子全書》第 6 冊。上海：上海古籍出版社；合肥：安徽教育出版社。

---（2002）。《朱子語類》（一）至（五），收入《朱子全書》第 14 至 18 冊。上海：上海古籍出版社；合肥：安徽教育出版社。

李光地（清）（1979）。《注解正蒙》。收入（清）紀昀等總纂，陸費墀總校，《欽定四庫全書》。台北：臺灣商務印書館。

---（1995）。《榕村語錄》。收入《榕村語錄・續榕村語錄》。北京：中華書局。

徐世昌等編纂（民）（2010）。《清儒學案》第二冊。北京：中華書局。

章梫纂（清）（1968）。〈任賢〉，《康熙政要》。台北：華文書局。

郭慶藩（清）（1991）。《莊子集釋》。台北：貫雅文化事業有限公司。

張伯行（清）（1965）。《正蒙注》。收入《張橫渠集》。台北：臺灣商務印書館。

張載（宋）（1983）。《張載集》。台北：漢京文化事業股份有限公司。

程顥、程頤著（宋）（1981）。《二程集》。北京：中華書局。

劉儓（明）（1995）。《新刊正蒙解》。收入《續修四庫全書》。上海：上海古籍出版社。

劉璣（明）（1966）。《正蒙會稿》。台北：臺灣商務印書館。

二、現代資料

丁為祥（1994）。《實踐與超越——王陽明哲學的詮釋、解析與評價》。西安：陝西人出版社。

史革新（2007）。《清代理學史》。中山：廣東教育出版社。

牟宗三（1981）。《心體與性體》。台北：正中書局。

杜保瑞（2005）。《北宋儒學》。台北：臺灣商務印書館。

---。〈對王陽明批評朱熹的理論反省〉，《國立臺灣大學哲學論評》，44：33-72。doi: 10.6276/NTUPR.2012.10.(44).02.

林樂昌（2012）。《正蒙合校集釋》。北京：中華書局。

胡元玲（2004）。《張載易學與道學：以《橫渠易說》及《正蒙》為主之探討》。台北：臺灣學生書局。

唐君毅（2004）。《中國哲學原論‧原教篇》。台北：臺灣學生書局。

楊向奎（1985）。《清儒學案新編（一）》。濟南：齊魯書社。

楊儒賓（2007）。〈檢證氣學——理學史脈絡下的觀點〉，《漢學研究》，25，1：247-281。

鄭宗義（2009）。《明清儒學轉型探析——從劉蕺山到戴東原（增訂版）》。香港：中文大學出版社。

劉述先（2012）。〈張載在宋代理學的地位重探〉，郭齊勇主編，《儒家文化研究》，4：1-20。北京：新華書店。

技藝與理解

蔡政宏*

摘　要

　　人是如何獲得「理解」？德性知識論的代表人物 Linda Zagzebski 長久來宣稱「理解的獲得來自於對技藝的掌握」。在本文中，作者闡釋 Zagzebski 的宣稱，並指出這宣稱的問題。透過對 Zagzebski 之宣稱的批判性檢視，作者在文後提供另一修改自 Zagzebski 的理解獲得觀點。

關鍵詞：技藝、技巧、解釋、識知、理解

---

*　蔡政宏，東吳大學哲學系副教授。

Technê and Understanding

Cheng-Hung Tsai[*]

Abstract

How can we acquire understanding? Linda Zagzebski has long claimed that understanding is acquired through, or arises from, mastering a particular practical technê. In this paper, I explicate Zagzebski's claim and argue that the claim is problematic. Based on a critical examination of Zagzebski's claim, I propose, in conclusion and in brief, a new claim regarding the acquisition of understanding.

Keywords: technê, skill, explanation, epistêmê, understanding

[*] Associate Professor, Department of Philosophy, Soochow University, Taipei, Taiwan.

Technê and Understanding*

Cheng-Hung Tsai

I. Introduction

Several contemporary epistemologists, particularly Linda Zagzebski (1996; 2001; 2009), Jonathan Kvanvig (2003), and Duncan Pritchard (2010), have urged philosophers to focus more on the very idea of understanding because, as they have argued, understanding is more valuable than knowledge.[1] These epistemologists not only give reasons for the shift of focus in epistemology but also offer their own accounts of understanding. Among these, Zagzebski's account of understanding (1996: Part I, Sect. 2.2; 2001; 2009: Ch. 1 and Ch. 6) deserves special attention due to its broader concern, which addresses the issues of the nature, possibility, and acquisition of

[1] For an overview of the recent literature on the value of understanding, see Stephen Grimm (2012), in which he examines the three main reasons that philosophers have had for thinking that understanding is more valuable than propositional knowledge: first, understanding is more transparent to the mind; second, understanding reflects or mirrors the world more profoundly; and third, understanding is a greater cognitive achievement.

understanding. Zagzebski identifies several features that distinguish understanding from knowledge; some features are concerned with the *nature* of understanding and others are concerned with the *acquisition* of understanding (a dimension that is often overlooked in the literature). In this paper, my focus is on Zagzebski's account of the acquisition of understanding,[2] especially her central claim that *understanding arises from mastering a technê*. Zagzebski has long argued for this claim, and I have been puzzled by her claim since it first published in 2001. However, no one has ever seriously scrutinized this claim. The claim deserves close scrutiny because if understanding, as epistemologists argue, is more valuable than knowledge and if we desire to have such a higher epistemic good, then a critical study of Zagzebski's account of the acquisition of understanding, which seems to be the only such account available in the epistemological literature, shows us what can and what cannot be a viable way to acquire understanding. The structure of this paper is as follows. In Section II, I explicate Zagzebski's claim. In Section III, I argue that the claim is problematic because the argument for it is either unsound or equivocal. In Section IV, based on the critical study of Zagzebski's account, I suggest a new claim regarding the acquisition of understanding.

---

[2] For critical reviews of Zagzebski's account of the *nature* of understanding, see, e.g., Grimm (2006) and Pritchard (2010: sec. 4.3).

II. Zagzebski's Account of the Acquisition of Understanding

Zagzebski's account of understanding appears mainly in her article "Recovering Understanding" (2001) and Chapter Six of *On Epistemology* (2009). In "Recovering Understanding", Zagzebski identifies three features of understanding that distinguish it from knowledge:

> [T]hree features of understanding: [i] It is acquired through mastering a *technê*; [ii] its object is not a discrete proposition but involves the grasp of part/whole relations; and [iii] it involves representing some portion of the world non-propositionally. (2001: 242)

In *On Epistemology*, Zagzebski also mentions three features of understanding. The first two of the features are the same as [i] and [ii] above; the third feature is newly added but is not used to replace [iii], which is still endorsed in *On Epistemology*. Here are Zagzebski's characterizations of the features:

> [The first] is that understanding is connected with learning an art or skill, a *technê*. One gains understanding by knowing how to do something well …… This leads to the second idea, which is that understanding is not directed toward a discrete proposition, but involves grasping relations of parts to other parts and perhaps the relation of parts to a whole. …… There is a third feature of understanding that distinguishes it from propositional knowledge ……: Knowledge can

be acquired by testimony, whereas understanding cannot. (2009: 144-145)

I summarize these four features of understanding as follows (the names for each feature have been added by me):

(i) *The Practical Feature*: Understanding is connected with or acquired through mastering a technê;[3]

(ii) *The Holistic Feature*: Understanding involves grasping part/whole relations;[4]

(iii) *The Non-propositional Feature*: Understanding involves representing some portion of the world non-propositionally;[5]

---

[3] A precautionary note should be added here: Zagzebski acknowledges that "[s]ome instances of understanding are so easy that they require nothing more than simple past experience — for example, understanding a stop sign in the United States" (2009: 144). She calls such instances of understanding "easy understanding". Zagzebski is not concerned with easy understanding because she suggests that "the more interesting and significant examples of understanding are connected with skills" (2009: 144). The present paper is not concerned with whether *all* understanding is connected with a skill but with how understanding is connected with a skill.

[4] This feature is endorsed by many contemporary epistemologists, but the terminologies they use in charactering the feature may differ. According to Stephen Grimm, "In the case of understanding, the objects would be something along the lines of 'structures' (Linda Zagzebski), or 'systems' (Julius Moravcsik [1979]), or 'information chunks' (Jonathan Kvanvig [2003]), or 'dependency relations' (Jaegwon Kim [1994] and [Grimm 2006]). While these descriptions differ in various ways, if there is a common idea here it seems to be that understanding is directed at a complex of some kind — in particular, at *a complex with parts or elements that depend upon, and relate to, one another*, and that the mind grasps or apprehends when it understands" (emphasis added, 2012: 105). For Zagzebski, such part/whole relations "can be spatial, such as the relative location of sites in a city, and they can be temporal, as in a musical composition. An important kind of relation is that of cause to effect, or more generally, what Stephen Grimm calls dependency relations" (2009: 144).

[5] This is because, for Zagzebski, propositional structure does not exhaust the structure of reality. For instance, art, music, maps, graphs, diagrams, and causal nexus are portions of the world, and their structures are non-propositional. An understanding or state of comprehension of the non-propositional structures of reality can be acquired, according to the Practical Feature, through mastering a technê. Thus, Zagzebski says that "The *technai* of art, music, and literature can produce a state [of comprehension] that has epistemic value" (2001: 243).

(iv) *The Non-testimonial Feature*: Understanding cannot be acquired by testimony.

These four features can be classified into two groups. Features (ii) and (iii) are grouped together because they concern the nature of understanding. These features as a whole tell us what understanding is: understanding is the state of grasping the part/whole relations of the non-propositional structure of reality. Features (i) and (iv) are grouped together because they concern the acquisition of understanding. Further, (iv) can be treated as being derived from (i) because *if* understanding does involve mastering a technê and *if* a technê cannot be acquired by testimony or instruction alone (but by practice), then it follows that understanding cannot be acquired by testimony. Because my concern in this paper is the acquisition of understanding and because (i) is theoretically more fundamental than (iv), I shall focus on (i).

What does Zagzebski mean by the Practical Feature (I shall use the term as if it were Zagzebski's term)? First, let us see why Zagzebski thinks that the Practical Feature can be used to distinguish understanding from propositional knowledge:

> [U]nderstanding (*epistêmê*) … **has something to do with** *technê* － practical human arts or skills. … Understanding is a cognitive state that **arises from** *technê*, and since *technê* includes certain practical activities that are by no means wholly cognitive, it follows that understanding … is a state that **arises from** practices that are not purely cognitive. (bold emphasis mine, 2001: 240)

Can Zagzebski's aim of distinguishing understanding from propositional knowledge be successfully achieved by proposing the claim that understanding has something to do with or arises from technê and that propositional knowledge does not? At this moment, it appears difficult to answer this question confidently because we need more information about how Zagzebski characterizes the key phrases in the claim, i.e., "having something to do with" or "arising from", to determine whether propositional knowledge has something to do with or arises from technê. Here are two possible cases that undermine Zagzebski's aim of distinguishing understanding from knowledge. First, it is possible that some propositional knowledge "has something to do with" technê. For example, according to intellectualism about knowing-how (or technê), knowing-how is a species of knowing-that. Intellectualism argues that an agent's *knowing how*, for instance, to ride a bicycle skillfully amounts to her *knowing that* such and such is the way for her to ride the bicycle, and the agent entertains the proposition about the way in which she does so under the practical mode of presentation (Stanley & Williamson, 2001; Stanley, 2011). Under the intellectualist account of knowing-how (or technê), whenever an agent possesses a practical technê, he must know or entertain a certain proposition. Here, I do not suggest that intellectualism is correct (for criticism, see Tsai, 2011a, 2011b) but that without any qualification propositional knowledge also "has something to do with" technê, that is, the former constitutes the latter. Second, it is possible that all propositional knowledge "arises from" technê. For example, according to strong anti-intellectualism, knowing-that is a species of knowing-how; in Stephen Hetherington's words, "knowledge that p is the ability — the knowledge-how — such as to

respond, to reply, to represent, or to reason accurately that *p*" (2011: 44-5). Again, I do not suggest that strong anti-intellectualism is correct but that Zagzebski's claim that understanding arises from technê needs to be better clarified so that her aim can be achieved.

The claim "understanding arises from mastering a technê" tends to be interpreted as "understanding arises *spontaneously and directly* from mastering a technê". That is, when one has mastered a technê, some sort of understanding naturally but not necessarily emerges, and it emerges directly from one's mastery of that technê. I think that the idea contained in this interpretation is possible and plausible within the well-known model of skill acquisition that has been developed by Hubert Dreyfus and Stuart Dreyfus (1986). According to Dreyfus and Dreyfus, a skill is acquired and developed through five stages: novice, advanced-beginner, competent, proficient, and expert. When an agent achieves the level of proficiency or above, he gains a sort of holistic understanding that "effortlessly occurs" (1986: 28).[6]

---

[6] One of the main targets of Dreyfus' phenomenological model of skill acquisition is cognitivism, according to which "*all* mental activity is cognitive — that perception, understanding, **learning** and action are all to be understood on the model of fact gathering, hypothesis information, inference making and problem solving" (bold emphasis mine, 1988: 100). For Dreyfus, "We must be prepared to abandon the traditional [cognitivist] view that a beginner starts with specific cases and, as he becomes more proficient, abstracts and interiorizes more and more sophisticated rules. It might turn out that skill acquisition moves in just the opposite direction: from abstract rules to particular cases" (1988: 102). Dreyfus' five-stage model of skill acquisition can be introduced as follows: (This introduction has been kept as short as possible. For a criticism of Dreyfus' model, see, e.g., Selinger & Crease, 2002).

　　　Stage 1: Novice. Regulative rules are indispensable in the first stage of skill acquisition, and they are usually given by an instructor. But such rules are special, that is, they are context-free. According to Dreyfus, "the instruction process begins with the instructor decomposing the task environment into context-free features which the beginner can recognize without benefit of experience. The beginner is then given [context-free] rules for determining actions on the basis of these feature" (1988: 102). Here, "context-free" is understood in the sense that "[e]lements of the situations to be treated as relevant are so clearly and objectively defined for the novice that they can be recognized without reference to the overall situation in which they occur" (Dreyfus &

To me, the above interpretation is the most possible reading of Zagzebski's claim. Additionally, the idea within the interpretation can be plausible because

---

Dreyfus, 1986: 21). Because it is the rules that make a skilled behavior possible at this stage (or in the context-free situation), I call the behavior "rule-based behavior".

Stage 2: Advanced Beginner. A novice's performance improves to the second level "only after the novice has considerable experience in coping with real situations" (Dreyfus & Dreyfus, 1986: 22). "As the novice gains experience actually coping with real situations, he begins to note …… perspicuous examples of meaningful additional components of the situation. After seeing a sufficient number of examples, the student learns to recognize them" (Dreyfus, 1988: 103). In this stage, the student recognizes not only context-free elements but also new, situational elements. What the student relies on in responding to situational elements is "instructional *maxims*" rather than "instructional rules". Dreyfus uses the term "maxim" to "differentiate this form of instruction from the first, where strict *rules* were given as to how to respond to context-free *features*" (1988: 103). Let us call skilled behavior performed at the second stage (or in the real situation) "maxim-focused behavior".

Stage 3: Competence. "With increasing experience, the number of features and aspects to be taken account of becomes overwhelming. To cope with this information explosion, the performer learns …… to adopt a hierarchical view of decision-making. By first choosing a plan, goal or perspective which organizes the situation and by then examining only the small set of features and aspects that he has learned are relevant given that plan, the performer can simplify and improve his performance" (Dreyfus, 1988: 103). Unlike the last two stages, the performer at the third stage encounters problems in achieving a goal. Applying strict rules and maxims are not sufficient to solve the problems; however, the *competent* performer devises a strategy to cope with the problems. I shall call skilled behavior performed at the third stage (or in the problematic situation) "strategy-focused behavior".

Stage 4: Proficiency. In Dreyfus' study, "the proficient performer will be deeply involved in his task and will be experiencing it from some specific perspective because of recent events. Because of the performer's perspective, certain features of the situation will stand out as salient and others will recede into the background and be ignored" (Dreyfus & Dreyfus, 1986: 28). The proficient performer has the ability of "holistic discrimination and association", that is, the ability "to intuitively respond to patterns without decomposing them into component features" (Dreyfus & Dreyfus, 1986: 28). I shall call skilled behavior performed at this stage "understanding-focused behavior".

Stage 5: Expertise. "An expert generally knows what to do based on mature and practiced understanding. When deeply involved in coping with his environment, he does not see problems in some detached way and work at solving them, nor does he worry about the future and devise plans. …… An expert's skill has become so much a part of him that he need be no more aware of it than he is of his own body" (Dreyfus & Dreyfus, 1986: 30). Let us call skilled behavior performed at this stage "intuition-focused behavior". Unlike the proficient performer, who "will still find himself *thinking analytically about what to do*" (emphasis added, Dreyfus & Dreyfus, 1986: 29), an expert "*sees intuitively what to do*" without applying rules and making inferences at all" (emphasis added, Dreyfus, 1988: 106).

Given the above construal of Dreyfus and Dreyfus' five-stage model, rules, maxims, strategies, understanding, and intuition respectively are key to determining and manifesting skilled behavior at the various stages. And, as we have seen, in one's acquiring a skill, the role of codified rules fades away gradually, and intuition ultimately takes over as one's skill improves.

it receives support from Dreyfus and Dreyfus's model of skill acquisition. However, I doubt that Zagzebski would accept this interpretation of her claim. The reason for this doubt lies in the following citation:

> [U]nderstanding is a state gained by learning an art or skill, a *technê*. One gains understanding by knowing how to do something well, and *this makes one a reliable person to consult in matters pertaining to the skill in question.* (emphasis added, Zagzebski, 2001: 241; 2009: 144)

Why does an agent become a reliable consultant regarding a technê when the agent gains understanding by knowing how to do something well or mastering the technê? It is not necessary for an expert of a particular technê to be a coach or consultant of the technê (cf. Dreyfus & Dreyfus, 1986). So, why does Zagzebski assert this?

The answer to this question lies in how ancient philosophers such as Plato and Aristotle use the word "epistêmê". Zagzebski finds that Plato scholar Gail Fine (1990) translates the word "epistêmê" in Plato as "knowledge". However, one should be careful to remember that the term "knowledge" here is not understood in the same way as it is generally understood in contemporary epistemology. Zagzebski notes that Fine "stresses that it [*epistêmê*] is a form of knowledge that is closely connected to understanding" (Zagzebski 2009: 143, fn.9; 2001: 238). This Greek conception of epistêmê or knowledge (actually and accurately, understanding) in Plato's philosophy is explained by Fine (and Zagzebski agrees) as follows:

> On the account [of Plato] I have proposed, one knows more to the
> extent that one can explain more; knowledge requires, not a vision,
> and not some special sort of certainty or infallibility, but sufficiently
> rich, mutually supporting, explanatory accounts. Knowledge, for
> Plato, does not proceed piecemeal; to know, one must master a whole
> field, by interrelating and explaining its diverse elements (Fine 1990:
> 114; quoted by Zagzebski in her work of 2009: 143)

Zagzebski endorses Fine's explanation of the conception of epistêmê or
understanding and rephrases the explanation in her own words:

> [O]ne does not understand a part of a field *without* the ability to
> explain its place within a much larger theoretical framework, and one
> acquires the ability to do that *by* mastering a skill. (emphasis added,
> 2009: 143-4)

Here are two examples that illustrate her explanation:

> One does not have *epistêmê* [understanding] of an astronomical fact
> *without* interrelating and explaining its relation to diverse elements
> within the field of astronomy, and one can do that only *by* mastering
> the *technê* of being an astronomer. (emphasis added, 2009: 143-144)

> [O]ne does not have *epistêmê* [understanding] of some feature of
> human psychology *without* the ability to explain how that feature fits
> into the larger framework of human psychology, and that *requires*

having mastered the *technê* of the psychologist. (emphasis added, 2009: 144)

In Zagzebski's explanation, an agent's understanding (*epistêmê*) requires his ability to explain or give an account of the technê, and this ability requires the agent's mastery of the technê. Here, Zagzebski introduces an ability unmentioned in our previous discussion of her work, that is, the explanatory ability. This is the ability that makes an agent a consultant of the technê when he gains an understanding by mastering it. Whether an expert of a technê must simultaneously be a consultant of the technê who is able to articulate his reasons for his skilled action is open to debate (for a defense of the view that expertise requires articulacy, see, e.g., Annas, 2011a, 2011b; for a criticism, see, e.g., Stichter, 2007). My purpose is not to settle the debate but to show that Zagzebski's view of the relationship between understanding and technê is not as direct as it might appear. Actually, understanding is *mediated* by the explanatory ability in her account of understanding (especially when it is explained in connection with the ancient notion of *epistêmê*). Thus, Zagzebski will not accept the aforementioned interpretation that "understanding arises spontaneously and directly from mastering a technê".

III. Does Understanding Arise from Mastering a Technê?

After explicating the relationship between understanding and technê by explicating the notion of *epistêmê*, I suggest interpreting Zagzebski's central claim as meaning that "understanding arises *indirectly* from mastering a

technê (through mediation with the explanatory ability)". Can the claim, under this interpretation, be used as a distinctive feature of understanding to differentiate understanding from propositional knowledge as Zagzebski intends? To answer this question fairly to Zagzebski, I shall clarify this interpretation further.

In my explication of Zagzebski's claim, there are two kinds of abilities (broadly construed to include ability, skill, and technê) related to understanding. The first kind of ability is practical ability (skill or technê), such as swimming, riding a bicycle, and playing a piano. The second kind of ability is the explanatory ability, which takes a particular practical ability as its object of explanation. To highlight the difference and relationship between these abilities, the first can be expressed as a "*first-order* practical ability" and the second as a "*second-order* explanatory ability". When characterizing understanding, Zagzebski always makes her point by claiming that understanding arises from mastering (the first-order) technê rather than claiming that understanding arises from (the second-order) explanatory ability. For example:

> [U]nderstanding *requires* the mastery of a [first-order] *technê*, you cannot give someone understanding *without* teaching them the *technê*. (emphasis added, 2009: 145)

A possible explanation of why Zagzebski characterizes the situation in this way － that is, to attribute priority to the first-order technê over the second-order explanatory ability － is that she thinks that an *essential* relation exists between understanding and the first-order technê, although she

does not deny the importance of the second-order explanatory ability. If so, her central claim can be clarified further as "understanding arises *essentially*, though indirectly, from mastering a technê".

However, this clarified claim is problematic because the argument for it is either unsound or equivocal. Let me start with the first problem. Assume that claiming that X arises essentially from Y amounts to claiming that X requires Y. (In her explanation of how understanding is possible, it appears that Zagzebski uses the phrases "arise (essentially) from", "require", and "is acquired by" interchangeably; she also uses expressions such as "one cannot gain understanding *without* …" and "one gains understanding *by* …" to characterize what is required for understanding; see, e.g., the above quotations from Zagzebski. I think that Zagzebski needs to give an explicit specification of what she means by *what-is-required*. I will come back to this soon.) Zagzebski's argument for her claim that "understanding arises (essentially) from mastering a technê" can be constructed as follows:

(P1) One's understanding of a part of a technê requires one's ability to explain the relation of the part to the whole technê.

(P2) One's ability to explain a part/whole relation in a technê requires one's mastery of the whole technê.

(C) Therefore, one's understanding of a part of a technê requires one's mastery of the whole technê.

My criticism of this argument lies in the notion "require" (or "arise from" if one prefers that phrase). Because this argument concerns the acquisition of

understanding, the notion "require" in (P1) is supposed to indicate an enabling condition[7] that informs us of what makes it possible to acquire understanding (rather than informing us about the necessary and sufficient condition for the concept of understanding). Therefore, (P1) amounts to saying that the possession (and exercise) of an ability to provide an explanation or account of the relation between a part of a technê and the whole technê is the means to achieving the end of acquiring an understanding of the part. An explanatory ability, which provides an explanation of the relation between a part and the whole, is something the possession and exercise of which *enables* one to gain an understanding of a part of the whole of which it is a part. When an explanatory ability is successfully exercised, a part-whole explanation, i.e., understanding, is achieved. The notion "require" in (P2) must be used in the same way as in (P1), such that the conclusion, (C), can be uncontroversially derived. Therefore, (P2) is intended to indicate that the possession (and exercise) of a particular technê is the means to achieving the end of acquiring the ability to provide an explanation of the relation between a part of a technê and the whole technê. However, this appears problematic because a technê is not something the possession and exercise of which *enables* one to acquire an explanatory ability. When one successfully exercises a technê such as archery, what is achieved is hitting the target rather than acquiring an explanatory ability. Because (P2) is false, the argument is unsound.

One might respond to the above criticism by saying that the truth of (P2) can be secured. Such a defender might admit that the mastery of a technê is

---

7　For a detailed account of enabling conditions, see especially Cassam (2007).

not an enabling condition for acquiring an explanatory ability but insists that the ability still "requires" a technê. This is because, the defender continues, the explanatory ability requires a technê as its object of explanation. The ability to explain a part of a technê and the whole technê cannot be the very ability *per se* if there is no object － here, a technê － for it to explain. In sum, mastery of a technê is a *pre-condition* for exercising the explanatory ability. Therefore, (P2) holds, as was the original desire. Although this defense might make (P2) true, it simultaneously makes the argument commit the fallacy of equivocation because the key word "requires" in (P1) and (P2) is used with different meanings: the former refers to an enabling condition, and the latter refers to a pre-condition. The conclusion (C) is problematic because we have no clear idea of what it really says and what reasons support it.

Thus far, I have proposed two possible interpretations of Zagzebski's claim that "understanding arises from mastering a technê". These are as follows: first, "understanding arises spontaneously and directly from mastering a technê", and second, "understanding arises essentially, though indirectly, from mastering a technê". The first interpretation of Zagzebski's claim is the most probable at first glance, and the claim in the interpretation can be plausible within Dreyfus's framework of skill acquisition. However, the second interpretation and not the first interpretation appropriately suits Zagzebski's text. However, the claim in the second interpretation is problematic because the argument for the claim is either unsound or equivocal.

IV. Concluding Remarks

Based on the aforementioned explication and criticisms, I suggest that Zagzebski should withdraw the claim that "understanding arises from mastering a technê" from her account of understanding. However, I do not think Zagzebski's concern with the origin of understanding is misdirected. Like other philosophers who emphasize the importance of the study of understanding, Zagzebski is concerned with the nature of understanding. However, her concern is broader than that of most others. The questions that she investigates include (Q1) "What is understanding?", (Q2) "Is understanding attainable?", and (Q3) "How do we get understanding?" (cf. Zagzebski, 2009: 8). The Practical Feature can be seen as Zagzebski's answer to (Q3), but this has been shown to be problematic.

I suggest answering (Q3) in the following way: *Understanding arises essentially from mastering the second-order explanatory ability, which takes the first-order practical ability as its object of explanation*.[8] Two abilities are mentioned in this claim. However, we must be cautious about what explanatory power these abilities have. It is only the second-order explanatory ability that explains where understanding comes from; the first-order practical ability provides nothing significant regarding the acquisition of understanding, it is just a pre-condition for the second-order explanatory ability.

To conclude, the aim of this paper is to investigate the relation between technê and understanding and in particular to do this by examining the

---

[8]   For a related discussion, see Tsai 2011b and 2014, where I develop the idea of the dual structure of practical expertise.

Practical Feature of understanding proposed by Zagzebski, which has been expressed by her as the claim that understanding arises from mastering a technê. On the negative side, I have argued that the claim is problematic. Therefore, the Practical Feature is not true. On the positive side, I have suggested that understanding arises essentially from the second-order explanatory ability, the objects of explanation of which are the first-order practical abilities. I hope that these considerations have provided a basis for a promising approach for the further study of the acquisition of understanding.

References

Annas, Julia (2011a). *Intelligent Virtue*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199228782.001.0001.

--- (2011b). "Practical Expertise." John Bengson & Marc A. Moffett (eds.). *Knowing How: Essays on Knowledge, Mind, and Action* (101-112). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195389364.003.0004.

Cassam, Quassim (2007). *The Possibility of Knowledge*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199208319.001.0001.

Dreyfus, Hubert (1988). "The Socratic and Platonic Basis of Cognitivism." *AI and Society*, 2: 99-112. doi: 10.1007/BF01891374.

Dreyfus, Hubert & Dreyfus, Stuart (1986). *Mind over Machine*. Oxford: Blackwell.

Fine, Gail (1990). "Knowledge and Belief in *Republic* v-vii." Stephen Everson (ed.). *Epistemology* (85-115). Cambridge: Cambridge University Press.

Grimm, Stephen (2006). "Is Understanding a Species of Knowledge?" *British Journal of Philosophy of Science*, 57: 515-535. doi: 10.1093/bjps/axl015.

--- (2012). "The Value of Understanding." *Philosophy Compass*, 7(2): 103-117. doi: 10.1111/j.1747-9991.2011.00460.x.

Hetherington, Stephen (2011). *How to Know: A Practicalist Conception of Knowledge*. Oxford: Wiley-Blackwell. doi: 10.1002/9781118083178.ch6.

Kim, Jaegwon (1994). "Explanatory Knowledge and Metaphysical Dependence." *Philosophical Issues*, 5: 51-69. doi: 10.2307/1522873.

Kvanvig, Jonathan (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511498909.009.

Moravscik, Julius (1979). "Understanding and Knowledge in Plato's Philosophy." *Neue Hefte für Philosophie*, 15: 53-69.

Pritchard, Duncan (2010). "Knowledge and Understanding." Duncan Pritchard, Alan Millar and Adrian Haddock, *The Nature and Value of Knowledge: Three Investigations* (1-88). Oxford: Oxford University Press.

Selinger, Evan & Crease, Robert (2002). "Dreyfus on Expertise: The Limits of Phenomenological Analysis." Evan Selinger & Robert Crease (eds.), *The Philosophy of Expertise* (213-245). New York: Columbia University Press.

Stanley, Jason (2011). *Know How*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199695362.001.0001.

Stanley, Jason & Williamson, Timothy (2001). "Knowing How." *The Journal of Philosophy*, 98(8): 411-444. doi: 10.2307/2678403.

Stichter, Matt (2007). "Ethical Expertise: The Skill Model of Virtue." *Ethical Theory and Moral Practice*, 10: 183-194. doi: 10.1007/s10677-006-9054-2.

Tsai, Cheng-hung (2011a). "Linguistic Know-How: The Limits of Intellectualism." *Theoria: A Swedish Journal of Philosophy*, 77: 71-86. doi: 10.1111/j.1755-2567.2010.01091.x.

--- (2011b). "The Metaepistemology of Knowing-How." *Phenomenology and the Cognitive Sciences*, 10: 541-556. doi: 10.1007/ s11097-011-9208-0.

--- (2014). "The Structure of Practical Expertise." *Philosophia*, in press. doi: 10.1007/s11406-013-9513-7.

Zagzebski, Linda (1996). *Virtues of the Mind*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139174763.

--- (2001). "Recovering Understanding." Matthias Steup (ed.). *Knowledge, Truth, and Duty* (235-251). Oxford: Oxford University Press. doi: 10.1093/0195128923.003.0015.

--- (2009). *On Epistemology*. Belmont, CA: Wadsworth.

非單調推論及預建：評洛特之反例

傅皓政[*]

摘　要

　　此篇論文旨在以萊特於 1980 年所提出的預建邏輯系統，消除洛特對 *AGM* 理論的單調性原則提出的反例。相較於古典邏輯，雖然 *AGM* 理論主張知識狀態具備均衡性，並建構基本函映的預設說明信念變遷，但是許多哲學家認為該理論有某些基本原則值得質疑，洛特即指出 *AGM* 理論主張原有知識狀態的某個信念弱化之後，並不會影響原知識狀態的單調性並不成立。然而，洛特的反例之所以成立，在於可能出現相同語句卻具有不同資訊價值的情況，不過，洛特也並未提出適當的方式說明處理非單調的信念變遷過程。因此，我在這篇論文中要以 *AGM* 理論結合預建邏輯的方法消除洛特提出的反例，亦即主張在 *AGM* 理論中加入預建邏輯的推論規則，可以適當地處理非單調的信念變遷過程。

關鍵詞：非單調推論、預建邏輯、*AGM* 理論、信念變遷

---

[*]　傅皓政，中國文化大學哲學系副教授。

Nonmonotonic Reasoning and Defaults:
On Rott's Counterexamples

Hao-Cheng Fu[*]

Abstract

This paper aims to dissolve Rott's counterexamples in terms of default logic proposed by Reiter in 1980. In his counterexample, Rott pointed out that some fundamental principles in *AGM* theory which meet the property of monotonicity are implausible because it seems unable to handle the processes of nonmonotonic reasoning which are common in our ordinary reasoning. In contrast to classical logic, *AGM* theory indeed developed a prominent way to deal with belief change, whereas many philosophers still casted doubts on the fundamental principles in *AGM* theory. For example, there might be some sentences in one's belief set which are of the same form but different in information value and *AGM* theory does not commit it. Nevertheless, Rott failed to suggest a plausible way to fix the defect of *AGM* theory, so I aim to propose a suitable way to dissolve the problem of the processes of nonmontonic reasoning in virtue of the association of *AGM* theory with default logic.

[*]   Associate Professor, Department of Philosophy, Chinese Culture University.

非單調推論及預建：評洛特之反例[*]

傅皓政

壹、前言

　　自 *AGM* 理論的原型於 1985 年由阿克隆等人（Carlos Alchourrón et. al）提出之後，其建構的信念變遷模型一時蔚為風潮，該理論模型被廣泛地運用在經濟、計算機理論等領域。隨著 *AGM* 理論日漸受到重視，對於 *AGM* 理論本身的預設與定理等可能產生的問題，也就受到當代知識論與邏輯學者們更多的關注。本篇論文的主旨則是探討洛特（Hans Rott）於 2004 年對 *AGM* 理論所提出的反例，他宣稱這個反例是對 *AGM* 理論的幾個基本原則的挑戰，或者說，主要的問題在於這些基本原則所具備的單調性（monotonicity）性質，[1]並不能用來處理日常信念變遷的情況，也就是說，洛特主張日常信念變遷的模型應該符合非單調（nonmonotonic）的特性。

---

[1] 單調性的一般形式，從集合論的觀點來看，假設 H 和 H' 為兩個集合，m 為應用在此集合的函映，函映 m 具備單調性，若且唯若，若 $H \subseteq H'$，則 $m（H）\subseteq m（H'）$。

　　根據洛特的觀察（Rott, 2001; 2004），在 *AGM* 理論中，單調性這個性質可以從理論本身的預設推導而得。舉例來說，以修改函映而言，如果某個語句 *B* 已經在 *K*A* 這個信念集合中，亦即 *B*∈*K*A*，那麼所有在 *K*A* 這個信念集合中的信念，都會在信念集合 *K* 對語句 *A* 與語句 *B* 的連言（*A*∧*B*）修改的信念集合 *K**（*A*∧*B*）中被保留下來，即 *K*A* ⊆ *K**（*A*∧*B*）。然而，洛特認為在處理日常信念變遷的情況時，單調性這個性質並不成立，由此，他建構了一個反例來說明日常信念變遷其實是非單調的，原因在於不同的語句可提供的證成理由不盡相同，所以會推導出不同的信念，[2]因此，即使對原有的信念集合中可導出的信念運用修改函映，也可能會損失某些原來信念集合中的資訊。基本上，我們可以同意洛特的反例的確反映某些信念變遷的情況，而且 *AGM* 理論也的確無法處理這些信念變遷的情況。

　　根據洛特的描述，在日常信念變遷過程中，當然有許多情況是非單調的，但是洛特並未提出適合用以描述這些情況的邏輯系統，為了能夠適當地描述信念變遷的非單調性，我嘗試引進萊特（Ray Reiter）於 1980 年提出的預建邏輯（default logic）說明洛特的反例。事實上，當我們對運用修改函映操作某個連言（conjunction）時，構成此連言的連言項（conjuncts）之間，彼此可能分別證成不同的結果，而且這些結果彼此可能是不一致的，而依據 *AGM* 理論的預設，信念變遷的過程必須同時考慮信念集合的均衡性（equilibrium），因此除了原有 *AGM* 理論的預設外，還需要一些能夠說明處理此類信念變遷的推論規則。根據這個結構，我將首先說明 *AGM* 理論的函映具備單調性，接著，我會介紹洛特的反例及其問題，然後說明預建邏輯的基本結構，最後我主張如果以 *AGM*

2　值得注意的是，洛特在這裡所說的推導並不是嚴格演繹邏輯系統的推論規則的推導結果，而是根據其信念集合的評估所加入的信念。

理論結合預建邏輯系統能夠不但能消除洛特的反例，也能夠適當地說明非單調的信念變遷過程。

貳、*AGM* 理論與單調性

　　AGM 理論的目標主要是建構知識狀態的信念變遷模型，該理論的基本想法為在我們接受某些新資訊時，知識狀態會產生一些改變，*AGM* 理論建構的模型，正是要描繪知識狀態改變的過程。按照 *AGM* 理論建構的模型，知識狀態 K^3 改變的基本函映有三種，分別為擴展（expansion）、修改（revision）、縮減（contraction）函映。而且，這三種基本函映可以彼此相互定義。[4]因此，為了簡單起見，我們可以僅以修改函映說明 *AGM* 理論具備單調性這個性質。首先，*AGM* 理論修改函映的基本預設如下：（Gärdenfors, 1988: 54-56）

> （**K∗1**）對任意的語句 *A* 和任意的信念集合 *K* 而言，*K∗A* 是一個信念集合。
>
> （**K∗2**）$A \in K*A$。
>
> （**K∗3**）$K*A \subseteq K+A$。
>
> （**K∗4**）如果 $\neg A \notin K$，則 $K+A \subseteq K*A$。
>
> （**K∗5**）*A* 是矛盾句時，若且唯若，$K*A = K_\perp$。

---

[3]　根據 Gärdenfors 的說法，有許多方式可以用以描繪知識狀態 *K*，例如以信念集合（belief set）描繪知識狀態時，其定義如下：語句集合 *K* 是信念集合，若且唯若，（i）⊥ 不是語句集合 *K* 的邏輯結果；以及（ii）如果 $K \vdash B$，則 $B \in K$。從上述定義可以了解，信念集合 *K* 是包含其所有的邏輯結果，也就是說，$K=Cn（K）$。除了信念集合的描繪方式，Gärdenfors 還提到 Ellis 的信念系統（belief systems）、可能世界模型、以或然率建構的貝氏模型等等，為了討論上的方便，本文並不提及這些模型之間的關係，而以信念集合的方式描繪知識狀態。

[4]　基本函映相互定義的形式可以參考雷外同一（Levi's identity）或哈波同一（Harper's identity）。（Gärdenfors, 1988: 69-71）

（**K**∗6）如果 *A* 和 *B* 在邏輯上等值，則 *K*A=K*B*。

除了（**K**∗1）到（**K**∗6）的基本預設之外，在 *AGM* 理論的修改函映中，也引入兩個複合信念變遷的預設，這兩個預設與單調性有著密切的關聯，其預設如下：

（**K**∗7）*K**（*A∧B*）⊆（*K*A*）+*B*。
（**K**∗8）如果¬*B*∉*K*A*，則（*K*A*）+*B*⊆*K**（*A∧B*）。

（**K**∗7）與（**K**∗8）都涉及所謂的複合變遷，即重複地運用函映形成新的信念集合。其中，（**K**∗7）的意思是說，如果對原信念集合 *K* 而言，將修改函映運用在語句（*A∧B*），那麼所得到新的信念集合 *K**（*A∧B*）會是對原信念集合先運用修改函映操作語句 *A*，再用擴展函映操作語句 *B* 所得到的信念集合（*K*A*）+*B* 的子集合。而（**K**∗8）則進一步宣稱，如果語句 *B* 和原信念集合運用修改函映操作語句 *A* 所得到的信念集合一致的話，那麼上述的兩個信念集合會是相等的。

接著說明這兩個預設與單調性之間的關聯。首先，根據 *AGM* 理論處理信念變遷的模型，對於知識狀態而言，基本上遵守兩個重要原則，第一個原則是在信念變遷的過程中，信念集合會維持所謂的均衡性，亦即信念彼此之間不會有衝突。也就是說，信念集合至少不會出現不一致的情況，或者出現違反機率計算的基本原則等等（Gärdenfors, 1988: 9-10）。第二個原則是接受新資訊的原則，在修改函映的基本預設中，（**K**∗2）的意思就是說新資訊是運用修改函映操作之後所得到新的信念集合中的元素。所以，對原信念集合 *K* 而言，如果我們運用修改函映操作語句（*A∧B*），那麼（*A∧B*）顯然會是新的信念集合中的元素，同樣地，如果僅對語句 *A* 操作，那麼 *A* 也會是在新的信念集合中被接受的信念。

　　考慮以下情況，如果語句 B 已然出現在原信念集合 K 運用修改函映操作語句 A 所產生的信念集合 $K*A$ 中，即 $B \in K*A$，那麼對原信念集合運用修改函映操作語句（$A \wedge B$）所得到新的信念集合 $K*(A \wedge B)$ 而言，$K*A$ 這個信念集合會是 $K*(A \wedge B)$ 的子集合。因此，洛特認為在 AGM 理論的建構下，（R）無疑是成立的：[5]

　　　　（R）如果語句 y 已經在信念集合 $K*x$ 之中，那麼 $K*x \subseteq K*(x \wedge y)$。

對於（R），我們可以分成幾個情況討論：

　　第一種是 $K*x$ 不一致的情況，此種情況又可細分兩種不同的情況：（i）原始信念集合 K 是不一致的，由於不一致的信念集合可導出任何信念，因此 x 當然也是信念集合 K 的邏輯結果，如此，當信念集合 K 對語句 x 運用修改函映，由於語句 x 已出現在信念集合 K 中，所以，並不會產生任何新的變化，也就是說，$K*x$ 其實就等於 K，同理可證，語句 $x \wedge y$ 也是信念集合 K 的邏輯結果，因此，$K*(x \wedge y)$ 也等於 K，所以，$K*x = K*(x \wedge y)$，形式（R）成立。（ii）原始信念集合 K 是一致的，而且 $K*x$ 是不一致的，在這個情況下，根據（**K**∗5），我們可以得到語句 x 是矛盾句，或者說語句 x 本身是不一致的。[6]在上述兩種情況下，可以發現由於語句 x 是連言語句（$x \wedge y$）的邏輯結果，因此語句 x 當然是 $K*(x \wedge y)$ 這個信念集合的元素之一，因此，既然 $K*x$ 是不一致的，那麼 $K*(x \wedge y)$ 當然是不一致的。從這部分的證明可以了解，不可能出現 $K*x$ 是不一致而 $K*$

---

[5] 形式（R）一般稱為累積單調（*Cumulative Monotony*）或者慎推單調（*Cautious Monotony*），洛特認為其一般形式可以描繪如下：如果 y 在 $Inf(x)$ 中，則 $Inf(x) \subseteq Inf(x \wedge y)$，這裡的 $Inf(x)$ 基本上是採取 AGM 理論信念集合的觀點，也就是指所有以 x 為前提可以推論得到的結論。詳見 Rott（2004: 227）。

[6] 嚴格來說，不一致這個性質是應用在信念集合上而不是用在某個語句上，所以，在記號上的要求應該是 $\{x\}$ 是不一致的，即 "$\{x\} \vdash$"，為了簡單起見，在這裡我直接以語句 x 代表 $\{x\}$ 的情況，接下來的表示法也都相同。

（$x \wedge y$）卻是一致的情況。回到（R）本身，既然不一致的信念集合可以
推導出所有的信念，那麼在 K^*x 不一致的情況下，語句 y 當然是其元素
之一，那麼（R）當然也成立無疑。

　　第二種情況則是信念集合 K^*x 與 K^*（$x \wedge y$）均為一致的情況，按照
（R）的形式，語句 y 在信念集合 K^*x 之中，也就表示語句 y 為 K^*x 的
邏輯結果，既然如此，因為 K^*x 是一致的，根據（**K*2**），可以得到語句
x 也是信念集合 K^*x 的元素之一，另外，既然語句 x 和語句 y 都是信念
集合 K^*x 的元素，那麼語句（$x \wedge y$）顯然也是 K^*x 的元素，因此，語句
（$x \wedge y$）是一致的。接著，根據上述第一種情況的證明，我們可以得到當
K^*x 是一致的時候，原信念集合 K 和語句 x 都必須是一致的，所以，我
們最後可以得到 K^*（$x \wedge y$）也是一致的。而且，由於語句 y 已出現在 K^*x
之中，因此，K^*（$x \wedge y$）並未增加什麼新資訊，也就是說 $K^*x = K^*$（$x \wedge y$），
在這個情況下，（R）當然是成立的。

　　第三種情況則是設想信念集合 K^*x 是一致的，但 K^*（$x \wedge y$）卻是不
一致的情況。在這個情況下，由於 $K^*x \subseteq K^*$（$x \wedge y$）一定是成立的，因此
（R）也一定成立，因為（R）的後件為真。但是，值得注意的是這個情
況卻不是 AGM 理論要考慮的情況，因為當我們考慮 K^*x 是一致的，但
K^*（$x \wedge y$）卻是不一致的的情況時，由於語句 x 是 K^*x 的元素，因此語
句 x 本身會是一致的。另外，基於前述的說明，由於原信念集合 K 也是
一致的，因此，我們可以得到導致 K^*（$x \wedge y$）不一致的原因顯然來自語
句 y 或者語句（$x \wedge y$）。讓我們再分成以下兩種情況討論：（ⅰ）如果語句 y
是不一致的，同時語句 y 又是信念集合 K^*（$x \wedge y$）的元素，那麼 K^*（$x \wedge y$）
當然是不一致的。不過，在這個情況下，（R）的前件為假，因為語句 y
本身不一致，但 K^*x 是一致的，所以語句 y 不可能出現在信念集合 K^*x
之中。（ⅱ）如果是語句（$x \wedge y$）不一致的情況，語句 y 不一致的情況已包

含在情況（i）中，故略之。考慮語句 y 是一致的，但是語句（$x \wedge y$）是不一致的，在這個情況下，表示語句 y 會推導出¬x，或者說¬x 會是語句 y 的邏輯結果，同樣地，在這個情況下，（R）的前件為假，因為 K^*x 是一致的而且 $x \in K^*x$，而雖然語句 y 是一致的，但由於語句 y 會推導出¬x，因此，如果語句 y 在 K^*x 之中，那麼語句 x 和語句¬x 會同時出現在信念集合 K^*x 中，所以，K^*x 會是不一致的，由此可證，在此情況下，（R）的前件為假。

叁、洛特的反例

　　接著，讓我介紹洛特的反例。[7]設想甲國際上市公司要應徵人員填補一個行銷正職職缺，假設你是公司內部行銷部人員，所以你對於誰會成為你的同事深感興趣。截至目前為止，你的資訊如下：（i）公司只預備錄取一位正職人員；（ii）有 4 位應徵者通過初選得到面試的機會，分別為 a、b、c 和 d。這四位應徵者的背景分析如下：a 在行銷方面的資歷相當傑出，但對國際貿易幾乎毫無涉獵，而 b 對於行銷方面的資歷雖未如 a 那麼好，但是略懂國際貿易的各項操作流程，c 則是相當好的國際貿易專家，但是在行銷方面則顯得一無是處。最後，你發現前來應徵的 d 是目前市場上炙手可熱的行銷之星，是眾家公司爭搶的行銷寵兒，從這些背景加以分析，你會認為公司定然會錄取 d，當然，由於公司的正式職缺只有一位，所以 a、b、c 自然都未得到這份工作。假設這是你的原始信念集合 K，並且以語句 A、B、C、D 分別代表 a、b、c、d 獲得這份工

---

[7]　此反例為按照 Rott 的反例所改寫的版本，原來 Rott 是以應徵哲學系專任教師職缺為例，請參考 Rott（2004: 229-232）。

作，在原始信念集合 K 中至少有幾個與其相關的信念 $\{A{\lor}B{\lor}C{\lor}D$、$\lnot A$、$\lnot B$、$\lnot C$、$D\}$，簡記為 $K=\{A{\lor}B{\lor}C{\lor}D$、$\lnot A$、$\lnot B$、$\lnot C$、$D\}$。[8]

　　然始料未及的是，就在公司進行面試之前，你發現 d 已被另一個更大型的乙跨國企業公司錄取為專職行銷人員，並隨即撤回甲國際上市公司的應徵案。因此，在面試開始之前，你的原信念集合已經產生改變，即 $K^{*}{\lnot}D$，但是 $K^{*}{\lnot}D$ 並非僅僅是在原信念集合 K 中加入${\lnot}D$ 及減去 D 而已，因為甲國際上市公司預定錄取一位正職職員，如果僅僅是加入${\lnot}D$ 及減去 D，那麼新的信念集合 H 會成為$\{A{\lor}B{\lor}C{\lor}D$、$\lnot A$、$\lnot B$、$\lnot C$、${\lnot}D\}$，這個信念集合 H 顯然是不一致的，由於在 AGM 理論中要求除非語句本身是不一致的，否則運用修改函映之後的信念集合應該是均衡的，既然語句${\lnot}D$ 是一致的，因此 $K^{*}{\lnot}D$ 並不會等同於上述的信念集合 H，如果要讓信念集合 $K^{*}{\lnot}D$ 符合均衡性的條件，那麼在信念集合 H 中的元素，至少需減去$\lnot A$、$\lnot B$ 或$\lnot C$ 其中之一，才能保持一致性。然而，到底是要減去$\lnot A$、$\lnot B$ 還是$\lnot C$ 呢？這正是洛特接下來所要敘述的場景，原則上，在沒有更新資訊的情況下，$\lnot A$、$\lnot B$ 及$\lnot C$ 都會變成是暫時擱置的信念，因此，$K^{*}{\lnot}D$ 中的相關信念至少會出現 $\{A{\lor}B{\lor}C{\lor}D$、${\lnot}D\}$，不過，除此之外，由於甲國際上市公司僅預備錄取一位正職人員，因此，除了上述的相關信念外，在 $K^{*}{\lnot}D$ 中還會出現 $\{A{\to}(\lnot B{\land}\lnot C)$、$B{\to}(\lnot A{\land}\lnot C)$、$C{\to}(\lnot A{\land}\lnot B)\}$ 等相關信念。接著，讓我分析洛特描繪的場景，當然，不同的場景意味著 $K^{*}{\lnot}D$ 接受不同的新資訊所產生的信念集合：

---

[8]　根據先前所說 AGM 理論對信念集合的定義，信念集合 K 是包含所有邏輯結果，亦即 $K{=}Cn$ (K)，因此信念集合 K 是包含所有邏輯結果，因此其元素是無限多個，因為即使原信念集合 K 是空集合，也會包含所有的套套語句（tautologies），但是為了討論上的簡單起見，當寫成 $K{=}\{A, B\}$ 時，並不是指 K 中僅有 A、B 兩個元素而已，而是包含所有的邏輯結果，也就是 $K{=}\{A, B, A{\land}B, A{\lor}B, A{\lor}C, …\}$。

場景（1）：設想面試主管在面試結束後，向你提到應該是 a 或 b 會
　　　　　被錄取，這個訊息提供了 $A \lor B$ 這個前提，洛特認為根據
　　　　　這個前提，你可以推論 a 會被錄取這個結果，因為 a 在
　　　　　行銷方面的資歷更符合這個正職人員的要求。根據洛特
　　　　　的描述，當我們對（$K^* \neg D$）運用修改函映加入（$A \lor B$）
　　　　　這個信念時，則新的信念集合（$K^* \neg D$）$*$（$A \lor B$）中會
　　　　　產生的相關信念為 $\{ A \lor B \lor C \lor D、A \to (\neg B \land \neg C)、B \to$
　　　　　$(\neg A \land \neg C)、C \to (\neg A \land \neg B)、A \lor B、A、\neg B、\neg C、\neg D \}$。

場景（2）：設想面試主管在面試之後，向你提到 a、b 或 c 三個人
　　　　　都有可能會被錄取，雖然這個訊息提供了 $A \lor B \lor C$ 這個
　　　　　前提，然而對你而言，這個訊息提供的想法和場景（1）
　　　　　的訊息內容很不相同，原因在於 c 只是在國際貿易方面
　　　　　的專家，竟然也會成為選項之一，可見面試主管的考量
　　　　　已不僅限於行銷方面，而是加入是否具有國際貿易的專
　　　　　長，在此情況下，你會認為最有可能被錄取的人會是 b，
　　　　　換言之，對你而言，根據運用修改函映加入（$A \lor B \lor C$）
　　　　　之後，新的信念集合（$K^* \neg D$）$*$（$A \lor B \lor C$）中的相關信
　　　　　念包括 $\{ A \lor B \lor C \lor D、A \to (\neg B \land \neg C)、B \to (\neg A \land \neg C)、$
　　　　　$C \to (\neg A \land \neg B)、A \lor B \lor C、\neg A、B、\neg C、\neg D \}$。[9]

場景（3）：設想面試主管在面試之後，向你提到最後被錄取的是
　　　　　c，雖然大感意外，但在這個情況下，很明顯地你會接

---

[9] 為了增加推論結果的強度，洛特在談論哲學系專任教職的例子時，還特別以比例的方式呈現
每個候選人條件的優劣情況，例如將研究成果以 100 分計算，假設 a 在形上學的成績是 97
分，邏輯研究成果為 0 分，而 b 在形上學的研究成績是 92 分，在邏輯方面為 50 分等。所以，
如果在僅考慮形上學研究成績的情況下，a 會得到這份工作，但是當邏輯研究成績也列入考
慮時，b 顯然會勝出，相關內容可參照 Rott（2004: 230），在這裡我們不妨以相同的比例看
待甲國際上市公司的這個情況。

受語句 C 為真，因此，新的信念集合（K*¬D）*C 中的
相關信念會包括 ｛A∨B∨C∨D、A→（¬B∧¬C）、B→
（¬A∧¬C）、C→（¬A∧¬B）、¬A、¬B、C、¬D｝。

　　在洛特的分析中，其實有個非常重要的關鍵步驟，讓我們從場景（2）
開始，在場景（2）中，由於面試主管提到 a、b 或 c 會被錄取，根據這
個訊息，你推論得到語句 B 為真。但根據 AGM 理論的邏輯結果封閉原
則，由於語句 B 出現在信念集合（K*¬D）*（A∨B∨C）中，而且語句
A∨B 是語句 B 的邏輯結果，因此語句 A∨B 會在信念集合（K*¬D）*
（A∨B∨C）之中。回到形式（R），洛特認為如果我們以（A∨B∨C）取代
形式（R）中出現的語句 x，而且以語句 A∨B 取代形式（R）中出現的語
句 y，上述場景就會是形式（R）的反例。因為根據上述的說法，形式（R）
的前件為真，因為語句 A∨B 會在信念集合（K*¬D）*（A∨B∨C）之中，
但是形式（R）的後件卻為假，因為（K*¬D）*（A∨B∨C）並不會是（K*¬D）
（（A∨B∨C）∧（A∨B））的子集合，洛特主張後者的情況等於（K¬D）
（A∨B）。所以，在前者信念集合（K¬D）*（A∨B∨C）之中，根據洛
特場景（2）的描述，會得到語句 B 和語句¬A；但是，在後者信念集合
K*（A∨B）之中，根據洛特對場景（1）的描述，卻會得到語句 A 和¬B。
因此，我們可以找到語句 B 和語句¬A 是前者信念集合（K*¬D）*（A∨B∨C）
的元素，卻不是後者（K*¬D）*（（A∨B∨C）∧（A∨B））的元素。基於
上述證明，洛特主張（R）並不成立，而由於從 AGM 理論的預設可以證
明（R）是成立的，因此，如果洛特的說法是對的，那麼從形式（R）不
成立，就可以證明 AGM 理論的預設顯然是有問題的。

　　然而，我認為洛特建構的反例並不具備足夠的說服力，關於其反例
的缺陷，可以從兩部分加以觀察：（a）資訊價值不對稱的問題；（b）推
論規則的問題。接下來，讓我分別陳述這兩個問題。

（*a*）資訊價值不對稱的問題：

　　在洛特的反例中，信念依據其來源的不同，資訊價值也顯然存在著不對稱的問題，也就是說，在不同信念集合中所出現相同形式的語句，可能具有不同的資訊價值。在場景（1）中，面試主管提到 *a* 或 *b* 會被錄取的時候，這個資訊讓你的知識狀態接受 $A \lor B$ 為真的信念，洛特聲稱根據這些資訊內容進行分析，你會得到語句 *A* 為真的結果，然而這個推論過程雖然合理但卻不是經過嚴格的演繹邏輯規則所推導而得。很顯然地，就古典邏輯系統而言，語句 *A* 並非 $A \lor B$ 的邏輯結果，但是根據 *a* 和 *b* 的資歷背景加以分析，卻能合理得到 *a* 會被錄取，也就是語句 *A* 為真的信念。但是，在分析 $(K* \neg D) * ((A \lor B \lor C) \land (A \lor B))$ 這個信念集合的時候，洛特提到語句 $A \lor B$ 已出現在信念集合 $(K* \neg D)*(A \lor B \lor C)$ 中，其理由為語句 *B* 會出現在該信念集合中，而且語句 $A \lor B$ 為語句 *B* 的邏輯結果，因此，語句 $A \lor B$ 會出現在信念集合 $(K* \neg D)*(A \lor B \lor C)$ 中。在這個論證中，語句 $A \lor B$ 的資訊價值顯然是大不相同的，當此資訊來自面試主管時，它所提供的資訊能夠合理推導出語句 *A* 為真的信念，相對地，由語句 *B* 推論得到語句 $A \lor B$ 卻沒有相同的資訊價值。因此，雖然我們可以同意洛特所言，語句 $A \lor B$ 的確會出現在 $(K* \neg D)*(A \lor B \lor C)$ 這個信念集合之中，但是這裡出現的語句 $A \lor B$ 卻沒有多餘的資訊價值，換言之，我們不可能從語句 *B* 推導而得的語句 $A \lor B$，再進一步推論得到語句 *A* 也為真。

（*b*）推論規則的問題：

　　依據洛特的想法所建構的信念集合 $(K* \neg D)*((A \lor B \lor C) \land (A \lor B))$，根據 *AGM* 理論修改函映的預設，當我們運用修改函映對信念集合 $K* \neg D$

操作語句（（$A \lor B \lor C$）∧（$A \lor B$））時，以（**K∗2**）可知語句（（$A \lor B \lor C$）∧（$A \lor B$））是新的信念集合（$K^* \neg D$）∗（（$A \lor B \lor C$）∧（$A \lor B$））的元素。然而，因為（$A \lor B \lor C$）和（$A \lor B$）均為語句（$A \lor B \lor C$）∧（$A \lor B$）的邏輯結果，根據場景（1）和場景（2），當（$A \lor B \lor C$）出現在信念集合（$K^* \neg D$）∗（（$A \lor B \lor C$）∧（$A \lor B$））之中，會推導出 b 被錄取，也就是語句 B 為真的信念；而（$A \lor B$）卻會推導出 a 被錄取，也就是語句 A 為真的信念。不過，由於甲國際上市公司僅錄取一位正職人員，所以當（$A \lor B \lor C$）出現在信念集合（$K^* \neg D$）∗（（$A \lor B \lor C$）∧（$A \lor B$））時，會同時推導得到語句 $\neg A$ 為真的結果，同樣地，語句（$A \lor B$）也會同時推導出語句 $\neg B$ 為真的信念，據此，既然語句 A 和語句 $\neg A$ 均出現在信念集合（$K^* \neg D$）∗（（$A \lor B \lor C$）∧（$A \lor B$））中，那麼該信念集合顯然是不一致的，也就是（$K^* \neg D$）∗（（$A \lor B \lor C$）∧（$A \lor B$））$=K_\perp$。另一方面，信念集合（$K^* \neg D$）∗（$A \lor B$）卻不會有相同的結果，因為該信念集合只會推導出語句 A 為真而且語句 $\neg B$ 為真的結果，也就是說，在信念變遷過程中，信念集合（$K^* \neg D$）∗（$A \lor B$）可以保持一致性，即（$K^* \neg D$）∗（$A \lor B$）$\neq K_\perp$。所以，（$K^* \neg D$）∗（$A \lor B$）$\neq$（$K^* \neg D$）∗（（$A \lor B \lor C$）∧（$A \lor B$））。當然，洛特將（$K^* \neg D$）∗（（$A \lor B \lor C$）∧（$A \lor B$））視為等於（$K^* \neg D$）∗（$A \lor B$）的說法是根據古典邏輯系統的推論規則而得到的結果，因為語句（$A \lor B \lor C$）是語句（$A \lor B$）的邏輯結果，因此（$A \lor B \lor C$）∧（$A \lor B$）和語句（$A \lor B$）是等值的。從 *AGM* 理論的觀點來看，在其修改函映的預設中，如果語句 x 和語句 y 的是等值的，那麼信念集合 K 運用修改函映操作之後所得到的信念集合也會是相等的，即根據（**K∗6**），K^*x 會等於 K^*y。

　　基於上述的分析，我們可以發現當相同語句具有不同的資訊價值時，信念變遷的過程會大不相同，但是洛特卻未明確指出這點，才令人感到混淆。舉例而言，當某人的信念集合 K^*A 面對語句（$A \lor B$）時，可

能出現兩種不同的情況：（i）如果接受語句（$A \lor B$）的原因，僅僅是因為語句（$A \lor B$）是語句 A 的邏輯結果，那麼接受語句（$A \lor B$），信念集合 $K*A$ 顯然不會有任何變化。（ii）如果某人接受語句（$A \lor B$）的原因是有其他資訊價值時，信念變遷的過程如下：先將語句 A 自 $K*A$ 移除，然後再以新的信念集合（（$K*A$）$-A$）加入新資訊成為調整過後的信念集合（（$K*A$）$-A$）$*$（$A \lor B$）。[10]事實上，這兩種不同的信念變遷過程所形成的信念集合相當不同。

　　讓我用明確的日常實例說明得更清楚些。假設某位檢察官依據其手上的證據認定甲是嫌疑犯，在這個情況下，檢察官當然會同意甲或乙是嫌疑犯的說法，因為甲是嫌疑犯的信念足以支持甲或乙是嫌疑犯的信念，也可以說，後者不過是前者的邏輯結果而已，不但如此，該名檢察官也會同意甲或丙是嫌疑犯，或者是甲或丁是嫌疑犯等等的信念，也就是說，既然是從甲是嫌疑犯這個信念推論得到的邏輯結果，該選言的後項是乙、丙還是丁對於信念狀態而言並沒有任何實質影響。可是，如果該名檢察官依據後來得到的證據，使得他對於甲是嫌疑犯的信念產生懷疑，因此他移除甲是嫌疑犯的信念，而進一步加入甲或乙是嫌疑犯的信念的話，那麼當他加入甲或乙是嫌疑犯這個信念時，並不是將甲或乙是嫌疑犯這個信念當作甲是嫌疑犯的邏輯結果而已，換言之，乙是嫌疑犯這個選項對信念狀態並非是毫無影響的。由這個例子可以顯示出面對相同語句時，如果資訊價值不同，那麼信念變遷過程也會大不相同。

---

[10]　這個新的信念狀態也會是一致的，證明如下：由於語句（$A \lor B$）是語句 A 的邏輯結果，所以語句（$A \lor B$）和信念狀態 $K*A$ 顯然是一致的，而自 $K*A$ 中移除語句 A 時，（（$K*A$）$-A$）是 $K*A$ 的子集合，因此，語句（$A \lor B$）和信念狀態（（$K*A$）$-A$）也會是一致的。

肆、預建邏輯的推論規則

　　對於洛特建構的反例而言，我們目前可以同意 *AGM* 理論的確無法處理非單調推論的信念變遷過程，其理由在於 *AGM* 理論一方面接受古典邏輯系統的推論規則，另一方面 *AGM* 理論無法處理相同語句卻具有不同資訊價值的情況。但是，洛特也並未提供適當的模型處理不同資訊價值的語句，針對這個問題，我認為萊特於 1980 年提出的預建邏輯（default logic）提供了一個相當好的想法，能夠處理洛特的反例，即處理非單調推論的信念變遷過程。在使用預建邏輯解析洛特的反例之前，我想先介紹預建邏輯系統的記號與規則，基本上，預建邏輯系統乃是一般熟知的古典邏輯系統加上預建推論規則所組成的，所謂預建推論規則形式如下：

$$\frac{\alpha : \beta_1, \ldots, \beta_m}{\gamma}$$

其中$\alpha, \beta_1, \ldots, \beta_m, \gamma$都是語句，$\alpha$為前提，$\beta_1, \ldots, \beta_m$則稱為一致條件或證成，$\gamma$則是推論結果，如果$\alpha$是可導出的（derivable），而且和$\beta_1, \ldots, \beta_m$均一致（或者說$\neg\beta_1, \ldots, \neg\beta_m$不是可導出的），那麼可以推導得到$\gamma$。

　　接下來，我們需要定義預建邏輯中的外延（extensions），所謂的外延是指原信念集合經過預建推論規則得到某些推論結果形成的信念集合，因此，此外延的定義如下：設（D, W）是個預建理論，其中 D 是預建推論規則，W 是由事實與透過非預建推論規則推論得到的結果形成的語句集合。而 E 是語句集合，E 就是原信念集合運用預建推論規則之後

形成的信念集合。依據上述預建推論規則，如果α屬於 E，且¬β₁, ..., ¬βₘ 均不屬於 E，那麼γ屬於 E。

讓我們以一個簡單的例子來說明如何建立預建邏輯的外延，設想(D, W)為某個預建邏輯的理論，其中 W = { p } 而且 D 有兩組預建推論規則，分別為：

$$\delta_1 = \frac{p : q}{r} \qquad \delta_2 = \frac{\text{true} : s}{\neg q}$$

要從這兩組預建推論規則建立外延，建立外延的過程要確保外延是封閉的（closed），所謂外延是封閉的意思是，所有出現在 D 中可以運用的預建推論規則都必須被運用。首先，從δ₁ 開始，由於 W = { p }，所以其前提 p 已出現在原信念狀態中，而且 q 的否定並未出現在原信念狀態中，因此根據預建邏輯建立外延的方式，我們可以將 r 加入，因此形成{ p, r }，接下來觀察δ₂，前提位置中出現 true 的意思就表示不設定前提，所以，在未出現¬s 的情況下，就可以將其結論¬q 加入外延中，因此，最後形成的外延為 E_1 = { $p, r, \neg q$ }。第二個建立外延的方式則是從δ₂ 開始，同樣地以考慮 W = { p } 為起點，因為沒有出現¬s 的緣故，所以可以將¬q 加入外延中，因此形成 { $p, \neg q$ }，接下來考慮δ₁，由於一致性條件 q 的否定¬q 已出現在外延中，所以，我們無法將其結論 r 加入外延中，因此，最後形成的外延為 E_2 = { $p, \neg q$ }。

不過，在上述的例子中，我們會發現第一個方式形成的外延有些問題，因為在第一個步驟時，能夠把 r 加入外延是因為¬q 並未出現，但是在第二個步驟中，卻因為¬s 未出現，所以又將¬q 加入外延中，而 r 和¬q 並不是一致的，因此，這個外延似乎不是我們想要得到的結果。基於這個問題，我們必須引進另一個關於建立外延的重要原則，也就是成功的

（successful）外延。為了反映成功的外延，我們必須同時記錄外延及其否定的部分，讓我們以 IN（E）表示進入外延的語句集合，而且以 OUT（E）是必須排除的語句所形成的集合，所以，成功的外延必須滿足 IN（E）$\cap$OUT（E）$=\varnothing$的條件[11]。其中，IN（E）當然就是依據預建推論規則將符合規則的結論依序加入語句所形成的集合，而 OUT（E）則是記錄我們運用的預建推論規則中一致性或證成條件的否定，也就是$\neg\beta_1$, ..., $\neg\beta_m$ 等。簡言之，透過預建邏輯的系統，我們希望建立封閉且成功的外延。

　　為了能夠更清楚地呈現外延的建立過程，我們可以利用樹狀圖來進一步說明上述的例子，在樹狀圖中，每個點左側代表加入外延的語句集合，即 IN（E）；而點的右側代表需要排除了語句集合，即 OUT（E）。如果已經運用所有能夠運用的預建推論規則之後，我們可以比較最後出現的 IN（E）和 OUT（E），如果 IN（E）$\cap$OUT（E）$=\varnothing$，那麼該分支建立的外延是成功的，否則就是失敗的。上述的例子樹狀圖如下：

$$\text{Th}(\{p\}) \cdot \varnothing$$

$$\text{Th}(\{p, r\}) \cdot \{\neg q\} \qquad \text{Th}(\{p, \neg q\}) \cdot \{\neg s\}$$

$$\text{Th}(\{p, r, \neg q\}) \cdot \{\neg q, \neg s\}$$
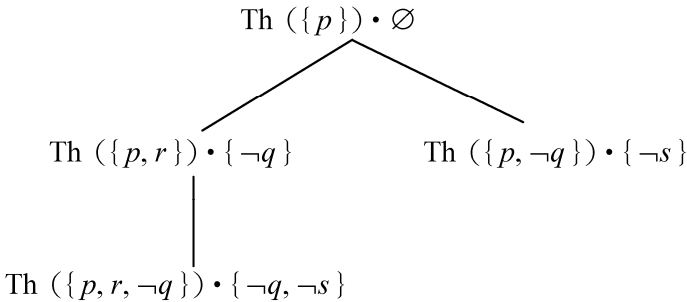
透過樹狀圖的分析，左分支是先運用預建推論規則δ_1，再運用δ_2 所建立的分支，仔細觀察左分支最後的點，左、右的語句集合中均出現$\neg q$，因

---

[11] 關於如何建立 IN（E）和 OUT（E）的建構方式，細節可參照 Antoniou（1999: 342-343）。

此，根據預建邏輯對外延的說明，此分支雖然是封閉的但不是成功的外延。接著，看到右分支的情況，由於¬q已出現在左側的語句集合，所以無法再運用預建推論規則δ₁，因此已經是封閉的分支，而且因為最後的點，其左、右側並未出現同樣的語句，所以同時是成功的外延。

伍、預建邏輯和洛特的反例

基本上，要解析洛特對不同場景的描繪，其實就是要指出即使是相同語句，也可能會有不同的資訊價值，要說明這些語句的資訊價值，可以藉由預建推論規則的方式完成，根據預建推論規則，可以說明為何可以從語句（$A \lor B \lor C$）推論得到語句 B 為真的結果，建構預建邏輯理論 $T=$（D, W）如下：

$$W= \{ A \to (\neg B \land \neg C), B \to (\neg A \land \neg C), C \to (\neg A \land \neg B), \neg D \}$$
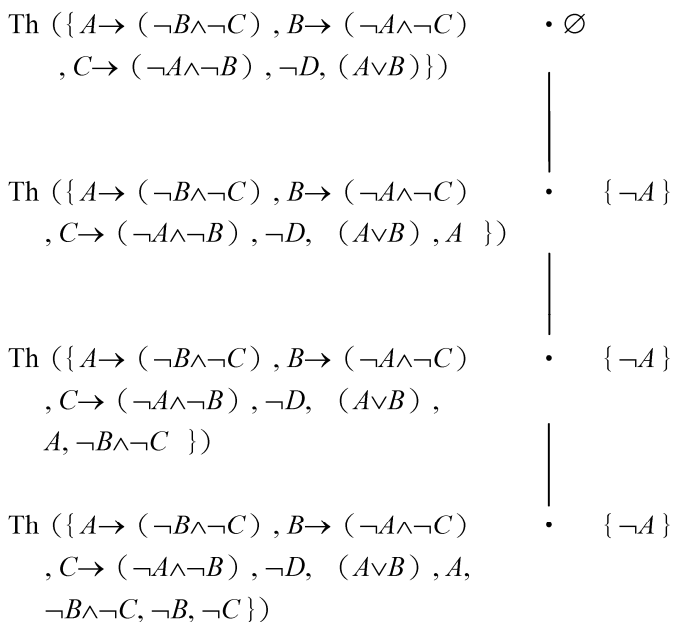
$$D= \{ \delta_1, \delta_2, \delta_3 \}$$

$$\delta_1 = \frac{A \lor B : A}{A} \qquad \delta_2 = \frac{A \lor B \lor C : B}{B} \qquad \delta_3 = \frac{true: C}{C}$$

接下來，讓我們進一步分析洛特的反例，按照洛特的說法，會造成信念集合產生問題的原因來自信念變遷過程中，並未維持信念集合的均衡性，以預建邏輯的觀點而言，就是作為信念集合並不是成功的外延。讓我以洛特建構反例的兩個主要路徑為例說明。

根據上述的分析，我們不難了解洛特反例主要的難題在於宣稱信念集合（$K*\neg D$）*（（$A \lor B \lor C$）∧（$A \lor B$））會等於信念集合（$K*\neg D$）*（$A \lor B$）

的說法，首先，讓我們以預建邏輯分析（$K^*\neg D$）*（$A\vee B$）的樹狀圖（1）如下：

樹狀圖（1）

Th（$\{A\rightarrow(\neg B\wedge\neg C)$, $B\rightarrow(\neg A\wedge\neg C)$　　　・∅
　　, $C\rightarrow(\neg A\wedge\neg B)$, $\neg D$, （$A\vee B$）$\}$)　　|
　　　　　　　　　　　　　　　　　　　　　　|
　　　　　　　　　　　　　　　　　　　　　　|

Th（$\{A\rightarrow(\neg B\wedge\neg C)$, $B\rightarrow(\neg A\wedge\neg C)$　　　・　$\{\neg A\}$
　　, $C\rightarrow(\neg A\wedge\neg B)$, $\neg D$, （$A\vee B$）, A　$\}$)　|
　　　　　　　　　　　　　　　　　　　　　　|
　　　　　　　　　　　　　　　　　　　　　　|

Th（$\{A\rightarrow(\neg B\wedge\neg C)$, $B\rightarrow(\neg A\wedge\neg C)$　　　・　$\{\neg A\}$
　　, $C\rightarrow(\neg A\wedge\neg B)$, $\neg D$, （$A\vee B$）,
　　$A, \neg B\wedge\neg C$　$\}$)　　　　　　　　　|

Th（$\{A\rightarrow(\neg B\wedge\neg C)$, $B\rightarrow(\neg A\wedge\neg C)$　　　・　$\{\neg A\}$
　　, $C\rightarrow(\neg A\wedge\neg B)$, $\neg D$, （$A\vee B$）, A,
　　$\neg B\wedge\neg C, \neg B, \neg C\}$)

根據樹狀圖（1）顯示，由於在其建構的外延中已出現$\neg B$和$\neg C$，因此，預建推論規則δ_2和δ_3均無法運用於建立新的外延，因為$\neg B$是δ_2中證成條件的否定，而$\neg C$是δ_3中證成條件的否定。再則，雖然語句（$A\vee B\vee C$）是語句$A\vee B$的邏輯結果，也會出現在信念集合（$K^*\neg D$）*（$A\vee B$）中，但是由於預建規則的限制，所以我們無法進一步經由接受語句（$A\vee B\vee C$）而得到語句B的結果，因此，不會出現不一致的情況。

　　同樣地，我們也可以建構（$K^* \neg D$）*（$A \lor B \lor C$）的外延，其樹狀圖
（2）如下：

樹狀圖（2）

Th（$\{A \to (\neg B \land \neg C)$, $B \to (\neg A \land \neg C)$　　　・　　$\varnothing$
　　, $C \to (\neg A \land \neg B)$, $\neg D$, （$A \lor B \lor C$）$\}$）

Th（$\{A \to (\neg B \land \neg C)$, $B \to (\neg A \land \neg C)$　　　・　　$\{\neg B\}$
　　, $C \to (\neg A \land \neg B)$, $\neg D$, （$A \lor B \lor C$）, $B\}$）

Th（$\{A \to (\neg B \land \neg C)$, $B \to (\neg A \land \neg C)$　　　・　　$\{\neg B\}$
　　, $C \to (\neg A \land \neg B)$, $\neg D$, （$A \lor B \lor C$）,
　　B, $\neg A \land \neg C$　$\}$）

Th（$\{A \to (\neg B \land \neg C)$, $B \to (\neg A \land \neg C)$　　　・　　$\{\neg B\}$
　　, $C \to (\neg A \land \neg B)$, $\neg D$, （$A \lor B \lor C$）, B,
　　$\neg A \land \neg C$, $\neg A$, $\neg C\}$）
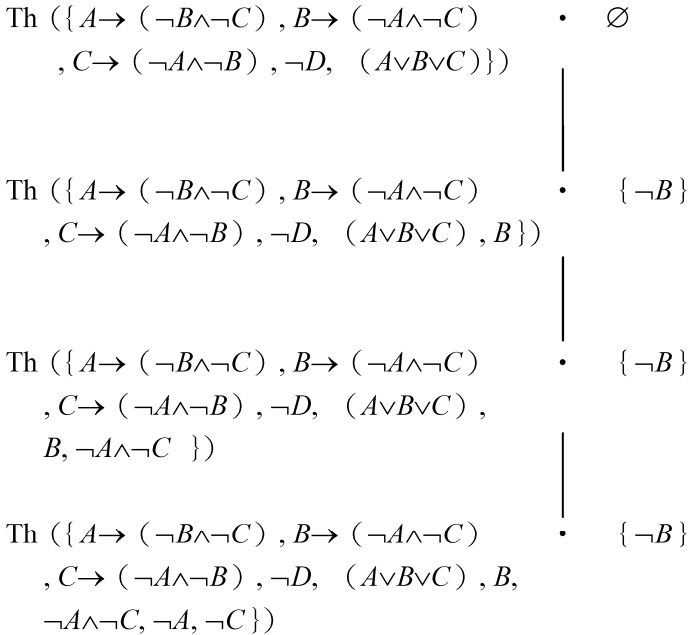
經由上述的樹狀圖（2），我們可以透過運用預建推論規則δ_2得到外延 $E=$
$\{A \to (\neg B \land \neg C)$, $B \to (\neg A \land \neg C)$, $C \to (\neg A \land \neg B)$, $\neg D$, （$A \lor B \lor C$）, B,
$\neg A \land \neg C$, $\neg A$, $\neg C\}$。接著，由於$\neg A$ 已經出現在 IN（E）的部分，因而無
法再運用預建推論規則δ_1 進一步得到新的外延，同樣地，由於$\neg C$ 也已
經出現在 IN（E）的部分，所以，也不能再運用預建推論規則δ_3 得到新
的外延。

　　根據我們所建立的預建理論，並沒有直接藉由語句($A \lor B \lor C$)∧($A \lor B$) 的資訊內容所建立的預建規則，因此，要分析（$K* \neg D$）*（（$A \lor B \lor C$）∧ （$A \lor B$）），可以得到其相關信念集合 {$A \rightarrow$（$\neg B \land \neg C$），$B \rightarrow$（$\neg A \land \neg C$）, $C \rightarrow$（$\neg A \land \neg B$），$\neg D$,（$A \lor B \lor C$）∧（$A \lor B$）}，由於（$A \lor B \lor C$）和（$A \lor B$） 均為（$A \lor B \lor C$）∧（$A \lor B$）的邏輯結果，因此，上述的相關信念集合也會 包含語句（$A \lor B \lor C$）和語句 $A \lor B$，因此，我們可以將上述相關信念集合 擴展為 {$A \rightarrow$（$\neg B \land \neg C$），$B \rightarrow$（$\neg A \land \neg C$），$C \rightarrow$（$\neg A \land \neg B$），$\neg D$,（$A \lor B \lor C$） ∧（$A \lor B$），（$A \lor B \lor C$），（$A \lor B$）}。接著我們可以看到，此時有兩個選項， 其一為先處理語句（$A \lor B \lor C$），即運用預建規則δ_2；另一個選項則是先處 理語句 $A \lor B$，亦即運用預建規則δ_1。如前述樹狀圖（1）及樹狀圖（2） 的分析，兩者都不會出現不一致的情況，所以，經由預建邏輯所建構的 外延，也能夠符合 *AGM* 理論對信念集合均衡性的要求。因此，我認為 洛特的批評部分是正確的，也就是 *AGM* 理論無法處理這些非單調推論 信念變遷的過程，但其反例牽涉到語句相同但資訊價值卻不同的情況， 洛特亦未能提出令人滿意的答案，不過，我認為如果在 *AGM* 理論中結 合建立預建理論，不但可以消除洛特的反例，也可以讓 *AGM* 理論具備 處理非單調推論信念變遷的情況。

陸、結論

　　洛特反例意圖在顯示即使是相同的語句，所具備的資訊價值可能也 不盡相同，但是 *AGM* 理論顯然無法處理這類的信念變遷過程。然而， 對 *AGM* 理論而言，其函映具備單調的性質也並非毫無爭議，至少在葛 登福斯的著作中，即討論過單調性並不成立的情況，他認為單調性原則 的一般形式如下：

　　　（**K*M**）如果 $K \subseteq H$，則 $K*A \subseteq H*A$。

設想以下情況，假設某個信念狀態 K，其中 $\neg A$ 不屬於 K 而 B 屬於 K，而且假設另一個信念狀態 H，H 為 $K+(B\to\neg A)$。由於 H 是對 K 運用擴展函映操作語句（$B\to\neg A$）形成的信念狀態，而且 B 屬於 H，所以，$\neg A$ 是信念狀態 H 的邏輯結果。接著，由於 $\neg A$ 不屬於 K，所以對信念狀態 K 運用修改函映操作語句 A 時，可以假設 B 仍然屬於 $K*A$。但是，如果對信念狀態 H 運用修改函映操作語句 A，則情況會大不相同，因為 $\neg A$ 是 B 和（$B\to\neg A$）的邏輯結果，所以，在 $H*A$ 中除了移除 $\neg A$ 以外，還必須至少移除 B 或（$B\to\neg A$），才能避免再導出 $\neg A$。當然，我們可以很容易想像在某些情況下，我們會移除 B 而不是（$B\to\neg A$），[12]所以，在這個情況下，雖然 $K \subseteq H$ 成立，但是 B 屬於 $K*A$ 卻不屬於 $H*A$，也就是說 $K*A \subseteq H*A$ 並不成立，亦即單調性原則並不成立。

　　不過，AGM 理論是否如洛特所言，對於已經出現在信念集合中的語句，運用修改函映操作該語句的結果，並不會使該信念集合產生任何變化？答案其實未必，亦即根據葛登福斯的觀點，此說法在 AGM 理論中並不見得成立，我們可以把這個想法的形式顯示如下：

　　　（**K*I**）如果 $\vdash A\to B$，那麼 $K*B \subseteq K*A$

葛登福斯同樣對（**K*I**）提出質疑，

　　　　設想 A 代表「這個火柴是濕的而且被摩擦」這個語句，而 B 代表「這個火柴被摩擦」。而 K 則是代表原來信念狀態，而且火柴未

---

[12] 這些情況不難想像，由於科學定律大多為條件句，因此在面臨事實證據與定律導出結果相左時，我們通常會傾向調整事實證據。在科學哲學的討論中，有許多關於特製性假設（*ad hoc*）的討論，或者關於保護帶（protect belt）的說法，都是設法保留科學定律的觀點。

> 被摩擦是已知的，也就是說，¬B∈K。我們有理由相信 K*B 會包
> 含火柴將會燃燒這個信念，但是 K*A 並不會包含這個信念。所
> 以，K*B 不是 K*A 的子集合。（Gärdenfors, 1988: 60）

當然，我們可以發現葛登福斯認為 AGM 理論的函映雖具備單調的特性，
但是如果面臨相關信念集合包含條件句的情況，尤其是反事實條件句
（counterfactuals）時，其信念變遷的過程的確不會符合單調的特性，但
洛特的反例焦點並不在條件句，而是邏輯結果關係。但是，經由上述的
分析，引進預建邏輯可以消除洛特的反例，同樣地，我認為葛登福斯考
慮的例子也同樣能夠以建構預建理論的方式加以解決。

References

Alchourrón, C., P. Gärdenfors & D. Makinson (1985). "On the Logic of Theory Change: Partial Meet Contraction Functions and Their Associated Revision Functions." *Journal of Symbolic Logic*, 50: 510-530. doi: 10.2307/2274239.

Antoniou, G. (1999). "A Tutorial on Default Logics." *ACM Computing Surveys*, Vol. 31, No. 3: 337-359. doi: 10.1145/344588.344602.

Arló Costa H. & A. P. Pedersen (2010). "Social Norms, Rational Choice and Belief Change." Erik J. Olsson & Sebastian Enqvist (eds.). *Belief Revision meets Philosophy of Science, volume 21 of Logic, Epistemology, and the Unity of Science*. Dordrecht, Heidelberg, London, New York: Springer. doi: 10.1007/978-90-481-9609-8_8.

Bochman, A. (2001). *A Logical Theory of Nonmonotonic Inference and Belief Change*. Berlin Heidelberg: Springer-Verlag.

Delgrande, James P., Abhaya C. Nayak & Maurice Pagnucco (2005). "Gricean Belief Change." *Studia Logica: An International Journal for Symbolic Logic*, Vol.79, No. 1: 97-113. doi: 10.1007/s11225-005-0496-7.

Gärdenfors P. (1988). *Knowledge in Flux: Modeling the Dynamic of Epistemic States*. Cambridge, MA: Bradford Books, MIT Press.

Reiter, R. (1980). "A Logic for Default Reasoning." *Artificial Intelligence*, 13: 81-132. doi: 10.1016/0004-3702(80)90014-4.

Rott, H. (2000). "Two Dogmas of Belief Revision." *Journal of Philosophy*, 97: 503-522. doi: 10.2307/2678489.

--- (2001). *Change, Choice and Inference*. Oxford: Oxford University Press.

--- (2004). "A Counterexample to Six Fundamental Principles of Belief Formation."
　　Synthese, 139: 225-240. doi: 10.1023/B:SYNT.0000024914.61925.5b.

知覺的反個體主義與視覺科學

梁益堉*

摘　要

　　本文從跨領域的角度探討視覺的根本性質，並以 Tyler Burge 的「知覺的反個體主義」（perceptual anti-individualism）為研究對象。根據這一理論，知覺狀態的本質乃是由知覺者與外在環境的互動關係而定。Burge 提出論證主張：視覺科學（vision science）預設了這個理論。本文反對這個觀點，並企圖從三方面來論證：「知覺的反個體主義」不是我們理解視覺的唯一理論選項。首先，我討論「體內恆定」（homeostasis）的概念，並指出這概念會使我們對 Burge 理論中的「知覺規範」（perceptual norms）產生質疑。第二，我以論證指出：許多視覺科學所研究的現象，可以不必預設 Burge 理論中的「正確性」（veridicality）和「單一表徵」（singular representation）也能得到解釋。第三，我討論一些有關視覺的科學理論並論證：許多視覺科學領域中的看法其實不支持 Burge 的理論。本文的結論是：「知覺的反個體主義」並不是瞭解視覺本質唯一可選的理論架構。

關鍵詞：知覺、視覺科學、知覺反個體主義

* 梁益堉，臺灣大學哲學系副教授。

Perceptual Anti-Individualism
and Vision Science

Caleb Liang[*]

Abstract

I discuss the nature of visual perception from an interdisciinary perspective. The target of investigation is Tyler Burge's theory of perceptual anti-individualism, according to which perceptual states constitutively depend on relations between perceivers and the external world. Burge argues that this theory is presupposed by vision science. My goal is to argue that perceptual anti-individualism is not the only theoretical choice. First, I consider the notion of homeostasis and suggest how it may cast doubt on the perceptual norms in Burge's theory. Second, I argue that many phenomena studied by vision science can be explained without positing Burge's notions of veridicality and singular representation. Third, I consider some empirical theories and argue that vision science does not uniquely favor Burge's theory. I conclude that perceptual anti-individualism is not the only framework for understanding visual perception.

Keywords: perception, vision science, perceptual anti-individualism

---

[*]　Associate Professor, Department of Philosophy, National Taiwan University.

Perceptual Anti-Individualism and Vision Science

Caleb Liang

I.Introduction

What is the most fundamental relationship between visual perception and the external world? One influential idea in vision science is that "visual perception is useful only if it is reasonably accurate" (Palmer, 1999). Tyler Burge has recently developed a philosophical theory that greatly articulates and substantiates this idea (2005, 2009, 2010). According to his theory of *perceptual anti-individualism*, the nature and individuation of perceptual states constitutively depend on relations, including causal relations, between perceivers and the environment. Drawing on diverse empirical resources, he argues for two important claims: first, perception delineates the lower border of representational mind and exhibits the most basic form of objectivity (2010: 10, 12). As he puts it, "Perception, representation, and objectivity begin together" (2010: 11). Second, perceptual anti-individualism is *presupposed* by vision science (2010: 87, 98-101). Burge contends that "perceptual anti-individualism provides the only acceptable framework for

understanding conditions under which perceptual representation is possible" (2005: 9). The goal of this paper is to argue against the second claim.[1]

To sharpen the focus, let me make two clarifications right away. First, my target is not the general form of anti-individualism[2] but its application to visual perception. The former is an abstract view about the individuation of mental states that, Burge holds, can be established without appealing to empirical research. I will not argue against the general anti-individualism, and I will not defend for any form of individualism, either. My concern is the latter, perceptual anti-individualism, which is a philosophical theory of vision. Since Burge claims that it is presupposed by vision science, it is open to empirical investigation whether this is indeed the case. Second, Burge says that even the thesis of perceptual anti-individualism itself, just stated above, is still abstract (2010: 87). As an abstract thesis, it can be compatible with many aspects of vision science. My concern is not with the abstract thesis. The issue to be addressed here is how well perceptual anti-individualism is filled in with details and supported by empirical studies. This is significant because whether vision science really presupposes Burge's theory ultimately depends on these considerations. As we will see, Burge articulates his theory with specific accounts of veridicality, perceptual norms, and elements of visual representation. These are the places where I will take issue with him.

---

[1]　In *Origins of Objectivity*, Burge forcefully criticizes various versions of Individual Representationalism, all of which, according to him, over-intellectualize the constitutive requirements of perception (cf. 2010: 12-22 and Part II). But in this paper I focus on his positive theory of perception.

[2]　Burge: "In its general form, anti-individualism is the claim that (A) the natures of many mental states constitutively depend on relations between a subject matter beyond the individual and the individual that has the mental states, where relevant relations help determine specific natures of those states" (2010: 61).

The aim of this paper is not to argue that perceptual anti-individualism is false. Rather, I argue for a more modest position, that is, this theory is not the only theoretical choice. Many aspects of vision science can be understood without assuming this theory. After presenting perceptual anti-individualism, three issues will be discussed. First, I consider the notion of homeostasis and point out how it may cast some doubt on the perceptual norms in Burge's theory. Second, I examine Burge's view of veridicality. According to this view, in order for a perception to be veridical, both the general and the singular elements of representational content have to be veridical. I argue that many phenomena studied by vision science can be explained without positing Burge's notion of singular representation. Third, I consider some empirical theories of vision and argue that none of them provides the sort of support that Burge requires. Vision science does not uniquely favor his theory. I conclude that perceptual anti-individualism is not the only framework for understanding how vision works.

II. Perceptual Anti-Individualism

Burge's theory is quite complex, but its core consists of two parts. The first consists of a set of *a priori* claims describing the constitutive nature of perception. The second part depicts an overview of vision science, which is meant to show how the science presupposes this theory. The *a priori* claims are the following.

(1) The constitutive nature of perceptual states depends on a systematic network of causal relations between instances of the environmental attributes and the individual. Burge considers this claim as a "necessary truth" (2010,

85). (2) Perceptual representation has veridicality conditions. Having veridicality conditions, according to Burge, is part of "what it is to *be* a perceptual state" (2010: 535).[3] (3) The representational function of perceptual states is to produce veridical representation. The success or failure of a perceptual state is to be evaluated according to whether it is an accurate representation of the objective world.[4]

(4) The representational content of perception has two elements. The *singular element* "functions fallibly to single out (refer to) perceived particulars" and is context-dependent (2010: 83, 381). This is to capture the idea that properties or kinds are never perceived in the abstract. An individual always perceives particular objects. The *general element* "functions fallibly to group or categorize particulars by attributing some indicated kind, property, or relation to them" (2010: 83, 380). This is to capture the other idea that perception necessarily represents particular objects as being a certain way.[5]

(5) "Perception is a capacity constitutively attributable to individuals" (2010: 369, 373, 536). It is individuals who perceive, not subsystems in the brain. This leads to the view that perception has biological functions in addition to the representational function. The biological functions of

---

[3] Burge: "a constitutively necessary condition of perceptual representation by an individual is that any such representation be associated with a background of some *veridical* perceptual representation" (2005, 1; cf. also 2010, 68).

[4] As pointed out by an anonymous reviewer, (2) and (3) in Burge's theory are not the same ideas. The link between them should not be taken for granted. For example, if (2) is plausible, it by no means implies that (3) is established. I do not address this potential issue in this paper. I thank the reviewer for this comment.

[5] Burge also holds that it is *a priori* that the representational content of perception constitutes a fallible egocentric perspective on such attributes and particulars (2010: 84, 401, 536). An individual always perceives an object from a given perspective, and the same object can be perceived from different perspectives.

perception contribute to survival, fitness and reproduction (2010: 301, 303). Burge emphasizes that the representational and biological functions of perception are two different kinds of functions; they are dissociable (2010: 302, 308, 411). A perception can be non-veridical but biologically useful. As he puts it, "Evolution does not care about veridicality. It does not select for veridicality *per se* (2010: 303)." Therefore, the representational function cannot be reduced to biological functions. Perception and representation are distinctive psychological kinds.[6]

The second part of the theory contains an overview that depicts three key aspects of vision science. The first is *the underdetermination problem of vision.* The primary goal of vision science, Burge claims, "is to explain how perceptual states that are of and as of the environment are formed from the immediate effects of proximal stimulation" (2010: 89). Consider the neural processing of object perception. Various patterns of light, reflected from external three-dimensional (3D) objects, strike the photoreceptors on the retina and form two-dimensional (2D) images of objects. These patterns of light are converted into neural impulses, carried by retinal ganglion cells. They travel through the lateral geniculate nucleus and enter into the visual cortex. The key is that, from 3D objects to 2D images, depth information about objects is forever lost. This creates a problem for the visual system. Different objects from different distances and orientations can project exactly the same 2D image on the retina. Theoretically, for any 2D retinal image there can be infinitely many possible distal causes. How does the visual

---

6　Regarding these *a priori* claims, I will leave (1) and (5) aside, even though I disagree with their idea of "constitutive". My criticisms in the next few sections will focus on (2) ~ (4).

system "figure out" which external object is the right one? This is called the "inverse problem" (Poggio et al., 1985; Palmer, 1999; Pizlo, 2008). Burge correctly characterizes this as a problem of underdetermination.[7]

Let me make a critical observation. There are actually two different problems involved here that need to be explained. One is that the actual *distal causes* are underdetermined with regard to the 2D retinal images. This is due to the fact that, with suitable orientations and distances, different external 3D objects can project identical 2D images on the retina. Some vision scientists call this the "ambiguity" problem (Purves & Lotto, 2003; Pizlo, 2008). The other is that the *perception* of an object can remain constant while proximal stimulation varies. For example, looking at an object from different perspectives, my retinal images are changing, but I still perceive the object as having the same shape and size. This is the problem called "perceptual constancy" (Palmer, 1999; Pizlo, 2008). In the case of shape, *shape ambiguity* is about how to identify the actual distal cause from indefinitely many potential causes on the basis of a single retinal shape. This does not involve changes in viewing perspectives. On the contrary, *shape constancy* essentially involves changes in viewing perspectives. It is about how to produce an invariant percept of shape, regardless of different 2D images, due to such changes. Since these are different problems, solving one does not imply solving the other.[8]

---

[7] Retinal information, Burge says, "significantly *underdetermine[s]* the distal causes of those registrations, hence the objects and properties that are represented in perception, hence representational content as of those objects and properties … The initial sensory registration of proximal stimulation in itself also underdetermines what perceptual representations the perceptual system will form" (2010: 90).

[8] Burge is aware of the difference between the ambiguity problem and the constancy problem. But

Facing the inverse problem (more precisely, the ambiguity problem), how does science explain 3D object perception? Burge argues that the mainstream of vision science decisively supports perceptual anti-individualism. The ways in which visual information is processed in the brain can be characterized as constrained or guided by what he calls *formation principles*.[9] These principles "privilege" or "bias" the neural process such that the underdetermined retinal inputs trigger a unique perception that (often but not always) represents the actual external object. The content of a perception is then determined by the operations of the formation principles embedded in the visual processing.

Obviously, the formation principles themselves require explanations. Perceptual anti-individualism plays an important role here. Burge says,

> In every case, formation principles … mirror basic facts in the broader physical environment. These are facts regarding spatial relations, natural forms of motion, the way light patterns tend to correlate with shadows and edges, the way surfaces tend to have unseen backsides, and so on. They mirror either environmental laws or deep environmental regularities that hold for the most part … So the natures of specific perceptual states are constitutively associated,

---

he thinks that both problems can be solved by what he later calls "formation principles" (2010: 92-94, 397-400).

[9] Burge: "The dominant scheme in the psychology of vision … is to explain a series of unconscious, largely automatic transformational processes that lead from registration of the array and spectral properties of light striking the retina to the formation of perceptions as of specific aspects of the distal environment … The transformations operate under certain principles that describe psychological laws or law-like patterns. These laws or law-like processes serve to *privilege* certain among the possible environmental causes over others … I call psychological principles that describe, in an explanatory way, these laws or law-like patterns *formation principles*" (2010: 92).

> via causal relations, with specific attributes, laws, and patterns in the environment. (2010: 98-99)

The idea is that the physical environment is not chaotic; it has various regularities. The formation principles do not simply come from nowhere. Rather, they stem from perceivers' long-term interactions with the world, and hence reflect those regularities of the environment. Burge holds that the formation principles can be explained "only by reference to the way in which patterns in the perceptual system's natural environment have molded the nature of the perceptual system and its perceptual states" (2010: 100). He discusses a few cases, such as convergence, lightness constancy, etc., and concludes that vision science is committed to perceptual anti-individualism.

The second aspect is a research tenet in the practice of vision science that Burge calls *the proximality principle* (2005: 22). He says:

> The formation of perceptual states depends causally, in any given instance, on registration of proximal stimulation. The same attributional kind of perceptual state, with the same attributional representational content, can be caused by the same type of registration of proximal stimulation, whether or not the perceptual state has perceptual *representata* － whether or not it is a perception *of* anything at all. (2010: 389)

According to this principle, the causal process of perceptions depends exclusively on proximal stimulation and visual processing in the brain. Let me add that there is an empirical justification for this tenet. The visual

system does not have infinite capacities to process innumerable distal objects. Rather, the visual system responds to similar objects or situations with similar *patterns* of processing. Vision science does not aim to explain particular cases; it studies patterns of interactions with the environment. It is very probable that the proximality principle captures how the visual system in the brain actually works. Hence, it is reasonable that the causal explanations provided by vision science are constrained by this principle.

The third aspect is *perceptual constancy*. Given the solution of the inverse problem and the proximality principle, the notion of objectivity in vision science is explained by the distinction between registration of sensory information and perceptual representation. Burge says,

> In effect, the transformation patterns systematically distinguish the merely proximal from the probably environmental … Specification of mind-independent and constitutively non-perspectival physical entities is separated out from the individual's sensory registration … Perceptual constancies are capacities for objectification … Objectivity is the product of separating what occurs on an individual's sensory surfaces from the significance of those stimulations for specific attributes and particulars in the broader environment. In this way, perception is the product of objectification. (2010: 398-400)

When I move towards a car, the visual information registered on my retina changes in a systematic way, and the 2D images gradually occupy a larger area in my visual field. However, the size of the car does not appear to change; it does not look to me as if it is getting bigger. This is "size

constancy." When I look at the Eiffel Tower and walk around it, the proximal sensory stimulations received by my visual system vary with respect to my pace, direction, and eye orientation. Yet the position of the Eiffel Tower does not seem to alter at all; it appears to me as being located in the same place. This is "position constancy."[10] According to Burge, objectivity is embedded in perceptual constancy studied by vision science. Perceptual constancy shows that what perception represents is a non-perspectival objective reality. For Burge, perceptual constancy is both necessary and sufficient for objective perceptual representation (2010: 413). Perceptual constancy draws the line between mere sensory information and perception, and it is this line that marks the beginning of perception, representation, and objectivity. Since perceptual constancy is also explained by formation principles, Burge concludes that it reveals how vision science supports his theory (2010: 346, 358, 365, 400).

III. Homeostasis and Norms of Perception

According to Burge's theory, the primary norm constitutively associated with perception is to "perceive things as they are — to form veridical perceptual representation" (2010: 312). To evaluate this claim, I would like to consider: How accurate should a perception be in order to be accurate *enough*? The rationale behind this question is that veridicality is a matter of degree. The notion of homeostasis will play a key role here. As researchers characterize it, homeostasis is "a dynamic and ongoing process comprising

---

[10] Perceptual constancy is pervasive; other forms of perceptual constancy include "color constancy," "lightness constancy," "shape constancy," "loudness constancy," etc.

many integrated mechanisms that maintain an optimal balance in the physiological condition of the body, for the purpose of survival. In mammals, these include autonomic, neuroendocrine and behavioral mechanisms" (Craig, 2003). Animals engage in various types of actions, such as eating, mating, navigating, predating, etc., for the purpose of maintaining homeostasis. For animals, the ecological environment is essentially characterized by its homeostatic values. In this regard, perceptual systems (and their biological functions) can be understood as part of the homeostatic system.

To see how this relates to our evaluation of perceptual anti-individualism, consider an important feature of the human visual system. That is, the neural-physiological structures of the human visual system are highly *specific* and *contingent*. Just to give a few examples: (i) Not all lights are visible. The wavelengths of visible light range roughly from 400-700 nanometers. (ii) The two kinds of photoreceptor cells, rods and cones, with roughly opposite sensitivities to light intensities, are not distributed evenly on the retina. The density of cones is high in the vicinity of the fovea and low in the periphery, which is opposite to the rods. The distributions of the three types of cone cells are also not even. (iii) The receptive fields of the retinal ganglion cells have a central-surround structure, which suggests that they are meant to represent *contrast* of light intensities rather than the exact spectrum of a particular light pattern. (iv) Vision acuity is high only when fixated by the fovea; peripheral vision is actually blurred. This is compensated for by various patterns of saccades. But during an ongoing saccade, the visual system does not take in any information at all. (v) During visual processing, a lot of visual information is dropped by the visual system. Only part of the

visual information gets into the range of visual attention and is stored in visual memory. Other contingent features include color contrast, filling-in, change-blindness, etc. (Palmer, 1999; Snowden et al., 2006; Baars & Gage, 2010).

The point is that, given the role of homeostasis, it is not obvious at all that these particular features are meant to help the visual system to "perceive things as they are." The visual system does not process every piece of information about the world that it receives, and it does not produce precise representations of everything in all their details that stand before our eyes. It seems that the perceptual states do not primarily *aim* to represent "things as they are" but to serve the needs of maintaining homeostasis. The visual system can function perfectly well to maintain homeostasis without fulfilling the primary norm proposed by Burge. The neural-physiological structures of vision do not unquestionably uphold the primary norm of Burge's theory.

Burge is aware of this potential objection. Some other norms in his theory attempt to respond to this worry. For example, he says: "A second natural representational norm constitutively associated with perception is to perceive as well as the perceptual system can, given its natural limitations, its input, and its environmental circumstances" (2010: 312). The visual system has various limitations, well-studied by vision science and recognized by his theory. He repeatedly emphasizes that the primary norm should be taken as an "idealization;" "perceptual representational contents when successful, are commonly only veridical within some range. So *approximate* veridicality is

what is often at issue" (2010: 535).[11] Adding this proviso makes perceptual anti-individualism more flexible to accommodate various non-ideal perceptual situations, and the primary norm remains central in the constitutive account of perception. As Burge comments, although the visual system is limited in many ways, "Still, veridicality is at the center of the natures and laws or law-like transformations that are central to perception and the subject matter of perceptual psychology" (2010: 535).

I do not think that the worry is really relieved by this proviso. First of all, since each specific limitation of the visual system is utterly contingent, making the proviso constitutive of the nature of perception seems a bit *ad hoc*. Moreover, given the contingent structures of the visual system, it is not quite right to say that strict veridicality is seldom achieved by perceptual representation. Compared with Burge's view, it seems more precise and adequate to say that there is *never* a perception that fully satisfies Burge's primary norm. Taking veridicality as idealization is not the best way to characterize the contingent features of the visual system. They can be better accommodated without appealing to Burge's primary norm and proviso.

An alternative way to consider the contingent features of the visual system, I suggest, is that *veridicality is constrained by homeostasis*. The question of how accurate is accurate enough is answered by homeostasis. I am not suggesting that veridicality should simply be replaced by homeostasis. I agree with Burge that the representational function of perception cannot be reduced to biological functions. My point is that the representational function of perception is not independent of biological functions, i.e. the former is

---

[11] This remark indicates that Burge agrees that veridicality is a matter of degree.

constrained by the latter. A successful perceptual representation does not need to be *more* accurate than homeostasis requires. A veridical perception is certainly helpful in maintaining homeostasis. Still, given the homeostatic constraint on veridicality, it is not mandatory to think that the representational content of perception must aim to enable animals to perceive things *as they are*. Not because the goal of achieving strict veridicality is hindered by various limitations of animals' visual systems, but because animals and their visual systems do not have to pursue this goal. The primary and secondary norms proposed by Burge are not the only theoretical choice for the purpose of understanding the nature of perception.

IV. Veridicality and Singular Representation

According to perceptual anti-individualism, it is constitutive of perceptual representation that it contains both general and singular elements. Perceptual states are produced by the visual system following formation principles, but the operations of formation principles are causally constrained by the proximality principle. Although what we perceive are particular objects, the visual system responds to similar objects or situations with similar patterns of processing. This is accommodated by the general element of perceptual representation. Or, as Burge states in another way, each perception has a "pattern-based" representational content (2005: 35).

Due to the proximality principle, it is possible that different perceptual contexts may trigger the same pattern of interactions, hence producing perceptions that are type-identical. For example, a subject may have a veridical perception of a particular object; on another occasion he may have a

veridical perception of a duplicate of the same type of object; on a third occasion he may have a matching hallucination of the object. The perceiver may be unable to distinguish between these situations because the same pattern-based representational content is produced. This means that, when the visual system is responding to the distal environment, it is responding to patterns or types of input (Burge, 2005: 5-6, 23-24). "The response to the input characterizes the distal environment as being of a certain kind" (2005: 6). So, as Burge characterizes it, the causal explanations provided by vision science "do not primarily explain particular events. They explain patterns, tendencies, general abilities, and so on" (2005: 32).

For Burge, a theory of perception must provide an account of how the veridicality of a perception is to be evaluated. It must be able to distinguish between veridical and non-veridical perception, and between seeing a particular object and seeing a numerically distinct duplicate. The general element, as constrained by the proximality principle, cannot reflect these differences. So Burge argues that perceptual representation must also contain a singular element to meet this requirement (2010: 389-390). It is the singular element that marks the differences among perceptions that are subjectively indistinguishable but with different veridicality values. For Burge, in order for a perception to be veridical, both the general and the singular elements of representational content have to be veridical (2010: 383). Let me call this a "full-hearted" notion of veridicality.

The issue that I want to raise is: Must we assume the "full-hearted" notion of veridicality in order to explain how vision works? I do not think so. It is my view that when a vision scientist says that "visual perception is

useful only if it is reasonably accurate" (Palmer, 1999: 6), the notion of accuracy in this remark can be understood as involving only the general element, not the singular element of perceptual content.[12] If so, from the standpoint of perceptual anti-individualism, the notion of veridicality in vision science would be only "half-hearted" in the sense that it takes into consideration only the veridicality of pattern-based representation. This would be unacceptable to Burge because he believes that the veridicality conditions of the subjectively indistinguishable situations mentioned above cannot be properly distinguished without positing the singular element. In the following I argue that there are various phenomena, seemingly supporting Burge's theory, can actually be explained by vision science without positing singular representation and the full-hearted notion of veridicality.

　　(1) *Perceptual constancy.* As Burge agrees, the visual processes that produce perceptual constancies are constrained by the proximality principle. He says: "exercises of the capacities [of perceptual constancies] are triggered even in cases where the proximal stimulations derive from no (environmental) objects of perception" (2010: 388). I fully agree with Burge on this point. A perceptual state can exhibit constancy without being veridical. A visual hallucination of an object can exhibit size or shape constancy, yet the singular element fails to refer to any object. However, this implies that perceptual constancy and veridicality are not the same. An empirical theory can explain perceptual constancy without embracing Burge's full-hearted notion of veridicality. As an essential feature of perception, perceptual constancy certainly reveals a type of objectivity. But this is a type of

---

[12]　To my knowledge, Palmer does not posit singular representation as formulated by Burge.

objectivity that can be understood without assuming Burge's theory. The distinction between sensory registration and perception drawn by perceptual constancy does not substantiate the kind of objectivity that is defined by Burge's notion of veridicality.

(2) *Visual illusions.* A defender of perceptual anti-individualism may contend that vision science has to posit the singular element in visual representation when it comes to explain visual illusions. For Burge, the formation principles of perception explain not only how the underdetermination problem is solved, but also "conditions in which perceptual systems yield misperception" (2010: 384). He says that "Failures of approximate veridicality ─ illusions ─ are explained primarily in terms of abnormal environmental conditions' producing proximal stimulations that would yield veridical representations under more normal conditions" (2010: 98). Since visual illusions are explained in terms of abnormal distal causes, Burge's notion of singular representation plays an essential role in vision science.

I disagree. Just because vision science appeals to distal causes to explain illusions, it does not follow that singular representation must be posited. Of course, what animals perceive are particular objects. But vision science does not aim to study *specific* particulars but their *properties* sharable with others. To see this point, let us consider: What would remain if all the pattern-based representations are removed from a perceptual state? The pattern-based visual representations of, say, a red flower, include representations of the color red, the shape and orientation of the flower, the distance in-between, etc. Suppose all these representations are removed. What remains would be a pure bearer or point of reference, which amounts to the singular element in

Burge's theory. Does vision science appeal to such a pure reference to explain illusions? I do not think so. Consider the case of Ames' Room, an interesting visual illusion in which the subject misperceives a trapezoidal room as rectangular, and misperceives an adult as absurdly smaller than a child in the room. The explanatory role of the distal causes, the room and the adult in it, is not to fix the pure references of what is seen. Rather, what explains the illusion is that, due to a manipulation of depth information, the *properties* of the distal causes, that is, the trapezoidal shape of the room and the actual height of the adult, are misrepresented by the pattern-based representations in the subject's perceptual state. The singular element in Burge's theory is unnecessary in this explanation.[13]

　　(3) *Multiple object tracking*. A defender of Burge's theory may draw on Zenon Pylyshyn's work on multiple object tracking (MOT) and his *visual index theory* (2003, 2007). The MOT experiment shows that normal subjects can visually track up to five randomly moving objects.[14] Pylyshyn explains

---

[13] The point here applies to many other illusions studied by vision science, e.g., the Müller-Lyer illusion, the Ponzo illusion, the Ebbinghaus illusion, etc. (Palmer, 1999: 324, 326). What about the distinction between a veridical perception of a particular object and a veridical perception of a numerically distinct duplicate? If, due to the same pattern-based representations, the subject mistakes the latter for the former, this can be regarded as a matter of thought being mistaken rather than false perception. What about the distinction between a veridical perception and a subjectively indistinguishable illusion? It is true that the notion of distal cause is required to draw this distinction. Still, it does not follow that vision science aims to explain how the pure reference of a perceptual state should be determined. At least in the cases discussed in this section, the explanatory powers of distal causes lie in their sharable properties rather than serving as pure references. The task of determining the veridicality of an indistinguishable illusion does not properly characterize what vision scientists do in their research.

[14] Pylyshyn: "In a typical MOT experiment, observers are shown a screen containing 8 simple identical figures (e.g., points, circles, squares, plus signs, figure eights) that move in unpredictable ways. … At the start of each trial, a subset of these objects is briefly rendered distinct (usually by flashing them on and off a few times). The observer's task is to keep track of this subset of objects. At some later time in the tracking trial (say 5 to 10 seconds into the trial) all the objects stop moving and the subject has to indicate (using a mouse pointing device) which objects were the

this by postulating a mechanism called "visual index" in the visual system. A crucial feature of this mechanism is that it enables us to track visual objects without representing any of their properties. Pylyshyn says that "the early visual system possesses a mechanism for detecting and tracking what I will refer to as 'primitive visual objects.' It does so by keeping track of them *as individuals* rather than as 'whatever is at location X' or 'whatever has property Y' " (2003: 180; cf. also 201, 206, 214; 2007: 38-39). In this sense, objects are "indexed directly" (2003: 202). A defender of Burge's view might consider the visual index theory as a strong case that vision science does posit singular representation in visual perception.

Unfortunately, if we look into Pylyshyn's theory, we will see that visual indexes are not considered as representations at all. He advocates a "conservative use of representations in theories", according to which "we should not postulate representations if no explanatory advantage is gained by such a postulate" (2007: 78).[15] This view is firmly embraced by Burge as well.[16] It is correct that, for Pylyshyn, no property is represented (or encoded) when an object is indexed by the visual system.[17] But this does not provide any defense for Burge's view. Two crucial points in Pylyshyn's theory are

---

targets. A large number of experiments have shown clearly that observers can indeed track up to 5 independently moving identical objects (i.e., objects that are indistinguishable by any property other than their historical continuity with the initially distinct objects) (2003: 223-224; cf. also 2007: 34-37).

[15] Pylyshyn distinguishes between information registrations and representation. The difference between the two lies in the fact that the former does not allow for the possibility of misrepresentation (2007: 74-75). If something plays a causal role in visual processing it does not follow that it plays this role by being represented (2007: 73).

[16] See Burge's criticisms of what he calls the deflationary conceptions of representation (2010: 292-308, especially 292, 299, 301).

[17] Pylyshyn: "no represented (or encoded) property is used in making the assignment of an index" (2003: 217; cf. also 2010: 81). Pylyshyn uses "represented" and "encoded" synonymously (2007: 68).

relevant here. First, as Pylyshyn says, "a sudden onset of a visual object may cause an index to be assigned without the assignment either being based on prior encoding of the event *as an onset* or itself carrying the information that an onset occurred" (2003: 218). This is to say that, as a causal event, a visual index is not itself encoded or represented. Second, Pylyshyn says that "Indexes, unlike codes, pick out things in the world to which they are related by a causal event, and they do not encode these things *as* something or other; indeed they do not encode them at all" (2003: 219). According to this point, even the indexed distal objects are not themselves encoded or represented. Hence, it is not the case that singular representation is required in the visual index theory. What one perceives are indeed particular objects, but it does not follow that it is mandatory to posit singular representation to explain multiple object tracking.

I have argued in this section that at least three types of visual phenomena － perceptual constancy, some visual illusions, and multiple object tracking － can be explained without appealing to singular representation or the full-hearted notion of veridicality. These phenomena, of course, do not cover all the areas in vision science. But they do challenge the claim that perceptual anti-individualism is the only framework for understanding how vision works.

V. Empirical Theories of Vision

In this section, I consider some empirical theories of vision and discuss how they may bear on perceptual anti-individualism. 3D object perception involves very complex processing. Most vision scientists would maintain that

we are still far from a full understanding of how it works. Currently, there are diverse accounts of the functions and mechanisms of visual perception. We will see that not all of them support Burge's position.[18]

(1) Let us first consider two major approaches to object recognition in cognitive science. The Recognition-by-Components Theory explains shape constancy by postulating 3D volumetric and viewpoint-independent units called *geons* (Biederman, 1987, 2007). Each object is analyzed and represented by the visual system as composed of a set of geons arranged in a certain way. The relations between geons are specified by a structural description using a viewpoint-invariant frame of reference, for example, a cylinder "on-top-of" a brick. A complete representation of an object is called a *geon structural description*, consisting of a set of geons together with a structural description of their relations. To perceive an object is for the visual system to generate a particular geon structural description for that object. It involves a series of processes: First, the edges of an object are extracted from the retinal images. Second, the non-accidental properties of the image are detected and parsed into regions. Based on this information, in the third stage, a particular set of geons is identified. Since geons are defined by non-accidental properties of retinal images, they explain why objects can be perceived as having the same shape regardless of changes in the retinal images. In the fourth stage, the geon structural description of an object is

---

[18] Burge regards the empirical account by Shepard (2001) and the Bayesian "ideal observer theory" by Geisler (2008) as supporting perceptual anti-individualism (Burge, 2010: 99). For critical evaluations of Shepard's view, cf. Kubovy and Epstein (2001) and Hatfield (2003). For critical comments on the Bayesian approach to perception, cf. (Purves & Lotto, 2011: 13-14) and (Purves et al., 2011, 15594).

compared with object representations stored in memory. The object of perception is recognized when there is a match.[19]

Another approach, the Multiple-View Theory, proposes a very different explanation of shape constancy, according to which object perception is essentially viewpoint-dependent and image-based. This theory consists of two main ideas. First, the visual system employs only 2D images to construct object perception. When one sees an object, a 2D image of the object is formed that depicts it from a particular viewpoint. Each 2D image represents various aspects of the object including shape, depth, color, texture, shading, etc., in a viewer-centered frame of reference (Tarr, 1995: 56). An ordinary object is represented by multiple 2D views that depict the object from various perspectives (Tarr & Bülthoff, 1995: 1495). Second, shape constancy is achieved by comparing 2D images on the retina with a collection of views already stored in memory. This is done by a set of transformation mechanisms (Bülthoff et al., 1994). When one perceives an object, the mechanisms bring the current retinal image into alignment with one of the

---

[19] The Recognition-by-Components Theory predicts that, when one visually identifies objects, the performance will not be affected by objects perceived from different perspectives. Biederman and Gerhardstein (1993) report several experiments that support this theory. In one experiment, 48 subjects were presented with a set of familiar objects, one at a time for 200 milliseconds (ms), followed by a mask for 500 ms (see Biederman & Gerhardstein, 1993: Figure 5). Then the set of objects was shown again for 100 ms, with some presented at the same orientation and some at different orientations (0°, 67.5°, and 135°). Although their orientations might change, the same parts of the objects would still be visible. Subjects were asked to name the objects as quickly and accurately as possible. Calculating the mean reaction time and error percentage on many trials, the results indicate that changing the orientations of the objects by rotation does not affect the subjects' performance. This suggests that object recognition is viewpoint invariant. The proponents of this theory also try to find supporting evidence in cognitive neuroscience. It is widely known that neurons in the macaque's inferior temporal cortex respond to shape properties (Tanaka, 1993). Hayworth and Biederman (2006) use an fMRI adaptation paradigm to argue that the lateral occipital complex is more sensitive to parts than to local features. They argue that these results fit nicely with the Recognition-by-Components Theory.

stored views by a set of mental transformations, such as rotations, translations, dilations, reflections, etc. (cf. Palmer, 1999: 364-365). The object is recognized when there is a match.[20]

For our purpose, the main point is that neither the Recognition-by-Components Theory nor the Multiple-View Theory lends obvious support to perceptual anti-individualism. As I argued above, a theory of vision can explain perceptual constancy without embracing Burge's full-hearted notion of veridicality. This means that Burge can only seek support from empirical theories that treat the ambiguity problem, rather than the constancy problem, as the central problem of vision. As we can see, both the Recognition-by-Components Theory and the Multiple-View Theory are about shape constancy rather than ambiguity. They can be understood and evaluated without assuming Burge's theory.

One possible defense of Burge's view is to argue that perceptual constancy is not totally independent of the ambiguity problem. Shape constancy cannot be fully explained by either of the two approaches above

---

[20] This approach predicts that the performance of object recognition will be affected according to how different the current view is from a familiar view in memory. Tarr (1995) conducted a series of psychophysical experiments to support this theory. In one experiment, 12 subjects were presented with seven left/right and front/back asymmetrical objects. After some training trials of viewing standard versions of the objects, both the standard versions and the mirror-reversed versions (produced by rotations of 130° around the *x*-, *y*-, or *z*-axis) were shown to the subjects, one at a time. They were asked to decide as quickly and accurately as possible whether the object was the standard or the mirror-reversed version of one of the objects that they had seen in the training trials. After calculating the mean reaction time and error percentage of many trials, Tarr reported the following findings: (1) the subjects' response time increased with the angular distance from the training viewpoint of seeing the standard versions of objects; (2) after some practice, performance became nearly equivalent at all familiar viewpoints; and (3) at unfamiliar viewpoints, response times increased with the angular distance from the nearest familiar viewpoint (Tarr, 1995, 64). Tarr argues that these findings strongly uphold the Multiple-View Theory. In addition, some neurophysiological data are considered as evidence for this theory. Logothetis *et al*. (1995) have reported that many neurons in monkey's inferior temporal cortex are sensitive to specific views of an object, and that different neurons encode different views. This, according to Tarr and Bülthoff (1998), provides evidence for multiple-view representation.

unless a solution to the problem of shape ambiguity is assumed. I believe that empirical support is required in order for this defense to begin to work, and I suspect it would not be easy to find.[21] Currently, as far as I know, empirical theories of object recognition explain only shape constancy. The defender of Burge's view demands that theories of object recognition *should* address ambiguity as well. It is not obvious whether most vision scientists would accept this demand. As we will see later in this section, even for those theories that accord the ambiguity problem a central place in vision research, not all of them concur with perceptual anti-individualism.

(2) Donald Hoffman (2009) advocates what he calls the "interface theory of perception," which employs the idea of a user interface from computer science. An icon on a computer screen has a particular color and shape and is associated with some stored file. But the icon's color and shape do not represent or reconstruct the "true" color or shape of the file. Computer files do not have any color or shape, and, as a user interface, the icon does not reconstruct anything. Hoffman holds that a user interface is useful precisely because it is not a reconstruction. He says: "The user interface is there to facilitate our interactions with the computer by hiding its causal and structural complexity" (2009: 154). A user interface is a convenient tool for specific purposes and nothing more.

Applying this idea to perception, Hoffman says that "Our perceptions are a species-specific user interface … to guide adaptive behavior in our

---

21   The support that this defense needs here, I think, is to find empirical theories that (1) explain both shape constancy and shape ambiguity, (2) formulate the two issues as connected in the way suggested by Burge, and (3) justify why shape constancy and shape ambiguity cannot be explained by pattern-based representations alone. As I see it, it is not easy to fulfill all of these requirements.

niche; accuracy of reconstruction is irrelevant" (2009: 154-155). Hoffman rejects what he calls the "principle of faithful depiction," the idea that the primary goal of perception is to provide veridical representation of the physical world (2009: 149). What perception does is not represent certain properties or categories of the objective world (2009: 153). Rather, "it is construction of a niche-specific, problem-specific, fitness-enhancing interface" (2009: 156). The aim of a vision theory is not to explain how veridical representations are produced by the visual system.

Like other empirical theories of vision, Hoffman's interface theory is controversial. It is not my goal to assess the explanatory power of the interface theory here. However, I will make two remarks with regard to our evaluations of Burge's theory. First, the conflicts between perceptual anti-individualism and the interface theory are obvious and serious. Proponents of both theories share an equal burden of defenses and criticisms. Since Burge maintains that perceptual anti-individualism provides the only framework within which vision science can be understood, a defender of his theory would need to establish that the interface theory is *in principle* incapable of explaining how the visual system solves the inverse problem. This is not an easy task. Until this is done, the interface theory remains a competitor.

Second, the dispute between Burge's view and the interface theory is very similar to the debate between scientific realism versus instrumentalism. The former claims － but the latter rejects － that the aim of vision science is to explain veridical representation of the world. If most researchers in applied sciences unreflectively take a realist stance about the world, this by

itself would not show that scientific realism is true. Likewise, employing the notion of veridicality in a philosophically naive way cannot be regarded as lending decisive support to perceptual anti-individualism. When vision scientists assume that we have veridical perceptions, it may just mean that they are naive realists. They do not, nor do they need to, consider whether their empirical research may help deal with certain philosophical issues. Therefore, in the practice of vision science, assuming that we often have veridical perceptions can just be a pragmatic or convenient choice, rather than a mandatory one. The interface theory illustrates that vision scientists do not have to presume perceptual anti-individualism to conduct empirical research and make sense of their work.

(3) Finally, let us consider a theory that is friendlier to Burge's account. Purves and Lotto (2003, 2011) maintain that the inverse problem is a central problem in vision science. They construe it as the ambiguity problem, and agree that it is a problem of underdetermination. However, they do not think that it can be solved by positing some *a priori* constraints or formation principles. Rather, they propose a purely empirical-statistical theory, which can be summarized as follows.

> [T]he visual system is not organized to generate a veridical representation of the physical world, but rather is a statistical reflection of visual history … By virtue of trial-and-error feedback over the eons about the success or failure of visually guided behavior in phylogeny and decades of ontogenetic experience, the visual brain simply responds to a stimulus with a pattern of neuronal activity

whose form has been thus determined by the probability distribution of what it has turned out to be in the past (i.e., by the *empirical significance* of the stimulus). (2003: 227-228)

In contrast to Burge, Purves and Lotto think that "visual percepts (and the corresponding activity of visual neurons and circuits) do not vary systematically with the physical measurements of objects or light stimuli as such" (2003: 15). The relations between vision and the world are purely contingent and statistical such that the visual system is not guided by formation principles that mirror environmental regularities. The solution of the ambiguity problem is gradually accumulated from the past. The ways in which the visual system responds to stimuli are partly trial-and-error. Then the system gradually learns from feedback over a long evolutionary history such that it acquires the capability of anticipation. That is, percepts produced by the visual system are, so to speak, fallible "predictions" of what is going to happen in the environment. This not only solves the ambiguity problem in theory, but also explains why animals are able to cope with the environment rapidly and efficiently.[22]

According to Purves and Lotto's account, the function of the visual system is not to produce veridical representations of the *current* physical environment. What is represented is the "probability distribution of the possible sources of the stimulus" (2003: 10). What animals see is whatever turned out to be the statistical majority of possible causes of visual stimuli *in the past*. Burge might think that this is compatible with his theory. But notice

---

[22] One of the anonymous reviewers has suggested me to take the anticipatory nature of perception into consideration. I am thankful for this useful comment.

that the statistical majority does not correspond to any particular object but to probability distribution. The feedback of trial-and-error and the statistical majority are understood not in terms of veridicality in Burge's sense, but in terms of biological functions. Also, visual illusions, according to this empirical-statistical theory, "are neither anomalies nor evidence of biological limitations or constraints, but simply the universal signature of this strategy of vision" (2003: 10). That is, there is no need to appeal to the perceptual norms suggested by Burge in order to understand vision.

VI. Conclusion

I have argued in this paper that various aspects of vision science can be understood without positing singular representation or the full-hearted notion of veridicality. Also, empirical theories of vision do not uniquely support perceptual anti-individualism. I conclude that, *pace* Burge, perceptual anti-individualism is not the only framework within which vision science can be understood. Let me make a final remark. Burge's perceptual anti-individualism has decisively elevated philosophical investigations of perception to a new and interdisciplinary level. I believe that what this theory has achieved is just a beginning, not the end. A lot more work can and must be done to deepen our understanding of the nature of perception.

References

Baars, J. & Gage N. (eds.) (2010). *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*, second edition. Oxford: Academic Press. doi: 10.1007/ s00381-010-1178-y.

Biederman, I. (1987). "Recognition-by-Components: A Theory of Human Image Understanding." *Psychological Review*, 94: 115-147. doi: 10.1037/0033-295X. 94.2.115.

--- (2007). "Recent Psychophysical and Neural Research in Shape Recognition." Osaka, N., Rentschler, I. & Biederman, I. (eds.). *Object Recognition, Attention, and Action* (Ch. 5, 71-88). Tokyo: Springer. doi: 10.1007/978-4-431-73019-4_6.

Biederman, I. & Gerhardstein P. (1993). "Recognizing Depth-rotated Objects: Evidence and Conditions for 3D Viewpoint Invariance." *Journal of Experimental Psychology: Human Perception and Performance*, 19: 1162-1182. doi: 10.1037/0096-1523.19.6.1162.

Bülthoff, H., Edelman, S. & Tarr, M. (1994). "How Are Three-dimensional Objects Represented in the Brain?" *CogSci Memo*, No. 5: 1-23. doi: 10.1093/cercor/ 5.3.247.

Burge, T. (2003). "Perceptual Entitlement." *Philosophy and Phenomenological Research*, LXVII, 3, November: 503-548. doi: 10.1111/j.1933-1592.2003. tb00307.x.

--- (2005). "Disjunctivism and Perceptual Psychology." *Philosophical Topics*, 33, 1: 1-78. doi: 10.5840/philtopics20053311.

--- (2009). "Perceptual Objectivity." *Philosophical Review*, 118, 3: 285-324. doi: 10.1215/00318108-2009-001.

--- (2010). *Origins of Objectivity*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199581405.001.0001.

Craig, A. D. (2003). "A New View of Pain as a Homeostatic Emotion." *Trends in Neuroscience*, 26, 6: 303-7. doi: 10.1016/S0166-2236(03)00123-1.

Geisler, W. S. (2008). "Visual Perception and the Statistical Properties of Natural Scenes." *Annual Review of Psychology*, 59: 167-192. doi: 10.1146/annurev. psych.58.110405.085632.

Hatfield, G. (2003). "Representation and Constraints: the Inverse Problem and the Structure of Visual Space." *Acta Psychologica*, 114: 355-378. doi: 10.1016/j.actpsy.2003.07.003.

Hayworth, K. & Biederman, I. (2006). "Neural Evidence for Intermediate Representations in Object Recognition." *Vision Research*, 46: 4024-4031. doi: 10.1016/j.visres.2006.07.015.

Hoffman, D. (2009). "The Interface Theory of Perception." Dickinson, S., Leonardis, A., Schiele, B. & Tarr, M. (eds.). *Object Categorization* (148-166). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511635465.009.

Kubovy, M. & Epstein, W. (2001). "Internalization: a Metaphor We Can Live Without." *Behavioral and Brain Sciences*, 24: 618-625. doi: 10.1017/ S0140525X01760086.

Logothetis, N., Pauls, J. & Poggio, T. (1995). "Shape Representation in the Inferior Temporal Cortex of Monkeys." *Current Biology*, 5: 552–563. doi: 10.1016/ S0960-9822(95)00108-4.

Palmer, S. (1999). *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press.

Pylyshyn, Z. (2003). *Seeing and Visualizing*. Cambridge, MA: MIT Press.

--- (2007). *Things and Places*. Cambridge, MA: MIT Press.

Pizlo, Z. (2008). *3D Shape: Its Unique Place in Visual Perception*. Cambridge, MA: MIT Press.

Poggio, T., Torre, V. & Koch, C. (1985). "Computational Vision and Regularization Theory." *Nature*, 317, September: 314-319. doi: 10.1038/317314a0.

Purves, D. & Lotto, R. (2003). *Why We See What We Do: An Empirical Theory of Vision*. Sunderland, MA: Sinauer Associates, Inc. doi: 10.5860/CHOICE.40-5803.

--- (2011). *Why We See What We Do Redux: A Wholly Empirical Theory of Vision*. Sunderland, MA: Sinauer Associates, Inc. doi: 10.5860/CHOICE.48-6268.

Purves, D., Wojtach, W. & Lotto, R. (2011). "Understanding Vision in Whole Empirical Terms." *PNAS*, 18, suppl. 3: 15588-15595. doi: 10.1073/pnas. 1012178108.

Shepard, R. (2001). "Perceptual-Cognitive Universals as Reflections of the World." *Behavioral and Brain Sciences*, 24: 581-601. doi: 10.1017/ S0140525X01000012.

Snowden, R., Thompson, P. & Troscianko, T. (2006). *Basic Vision: An Introduction to Visual Perception*. Oxford: Oxford University Press.

Tanaka, K. (1993). "Neural Mechanisms of Object Recognition." *Science*, 262: 685-688. doi: 10.1126/science.8235589.

Tarr, M. (1995). "Rotating Objects to Recognize Them: A Case Study of the Role of Viewpoint Dependency in the Recognition of Three-dimensional Objects." *Psychonomic Bulletin and Review*, 2: 55-82. doi: 10.3758/ BF03214412.

Tarr, M. & Bülthoff, H. (1995). "Is Human Object Recognition Better Described by Geon-structural-descriptions or by Multiple-views?" *Journal of Experimental Psychology: Human Perception and Performance*, 21, 6: 1494-1505. doi: 10.1037/0096-1523.21.6.1494

--- (1998). "Image-Based Object Recognition in Man, Monkey, and Machine." *Cognition*, 67, July: 1-20. doi: 10.1016/S0010-0277(98)00026-2.

編輯室報告

一、本期退稿率為 60%。

二、本刊下次出刊日期是 103 年 10 月，歡迎踴躍投稿。

三、投稿請依據本刊論文寫作格式。（公佈於臺灣大學哲學系網站 http://
　　www.philo.ntu.edu.tw）

《國立臺灣大學哲學論評》稿約

一、《國立臺灣大學哲學論評》刊登有關哲學研究之學術稿件，內容分下列各項：

    （一）論著

    （二）討論

    （三）書評

    （四）哲學界動態

二、本刊為半年刊，每年三月及十月出版，園地公開，歡迎海內外學界人士投稿。

三、本刊設編輯委員會，處理集稿、審稿、編印及其他出版相關事宜。來稿一律送請兩位學者專家審查，審查採雙匿名制；未獲採用者密退。

四、本刊接受中英文稿件，中文稿件以不超過二萬五千字（含註解）為原則，英文稿件以不超過 A4 紙四十頁（兩倍行高）為原則。

五、編輯委員會得就審查意見綜合討論議決，要求撰稿人對其稿件作適當之修訂。

六、稿件刊登後，每一撰稿人致贈本刊三本、抽印本三十份，不另致酬。

七、來稿以未在它處發表者為限。

八、來稿請標明中英文篇名、投稿者之中英文姓名，及五百字以內之中英文提要、五個以內中英文關鍵詞，並請務必按本刊「論文寫作格式」撰寫，以利作業。

九、來稿請附個人簡介，註明中英文姓名、服務機構、職稱、通訊地址、電話、傳真或電子郵件等聯絡資料，以便聯繫。

十、本刊著作者享有著作人格權，本刊則享有著作財產權；日後除著作者本人將其個人著作結集出版外，凡任何人任何目的之翻印、轉載（包括網路）、翻譯等皆須事先徵得本刊同意，始得為之。

十一、來稿請勿發生侵害第三人權利之情事。發表人須簽具聲明書，如有抄襲、重製或侵害等情形發生時，概由投稿者負擔法律責任，與本刊無關。

十二、來稿請包括文件稿三份（個人簡介另紙書寫）及以 word 檔儲存之電子檔案。來稿經刊登後，作者需簽署著作授權同意書二份，逕寄本刊編輯委員會，一份由本刊編輯委員會留存，一份由國立臺灣大學出版中心留存。

十三、來稿經本刊審查通過後，作者需同意授權本刊編輯委員會與國立臺灣大學出版中心，以及本刊編輯委員會或該中心所同意合作單位，依著作授權同意書之規定，進行紙本暨電子期刊之編輯、重製、翻譯、公開傳輸、收錄於資料庫等，提供使用者閱覽、下載與列印等，並授權該中心以紙本暨電子出版品方式發行，以利學術研究之促進。

十四、來稿郵寄地址為：（10617）臺北市大安區羅斯福路四段一號 臺灣大學哲學系《國立臺灣大學哲學論評》編輯委員會收或電子郵件信箱：ymlai@ntu.edu.tw。

《國立臺灣大學哲學論評》論文寫作格式

壹、正文

　一、文字書寫

　　(一) 文字書寫請以橫式寫作。

　　(二) 請用繁體字。

　二、章節層級

　　(一) 中文請按「壹、一、(一) 1. (1) a. (a)」等順序標示。

　　(二) 西文請按「I. A. (A) 1. (1) a. (a)」等順序標示。

　三、分段

　　(一) 每段首行前空兩格。

　　(二) 直接引原文時，短文可逕入內文。中文論文以「 」標出，西
　　　　 文論文以" "標出。

　　(三) 獨立引文自成一段落，每行前空三格。

　四、引文

　　(一) 請用新式標點符號，直接引文用冒號"："時，中文引號用「 」，
　　　　 西文引號用" "。

　　(二) 直接引文不用冒號"："時，中文引號用「 」，西文引號用" "。

　　(三) 引文中有引文，中文單引號「 」內之引文符號為雙引號『 』；
　　　　 西文雙引號" "內之引文符號為單引號' '。

　　(四) 獨立引文不加引號。

　五、書刊符號

　　(一) 書名、報刊名用《 》。

　　(二) 單篇論文及篇名用〈 〉。

　　(三) 書名與篇名連用時，省略篇名號，如《史記・項羽本紀》。

(四) 西文書名用斜體，篇名用" "。

六、附加原文

　(一) 引用西文人名、著作、專有名詞時，請盡量譯為中文。

　(二) 第一次引用西文人名、著作、專有名詞時以括號附加原文全
　　　名，其後引用時不需重複附加。

七、年代

　(一) 文中若有中國朝代名，請使用國字，並以括號附註西元年代，
　　　如明正德十六年（1521）。

　(二) 西元年代用阿拉伯數字，如董仲舒（176-104 B.C.）。

八、註釋

　(一) 於本文中註明出處，請按（作者，年代／縮寫：頁數）之格式
　　　加註，如：（Fine, 1994: 7-9），（Kant, KRV: 25(B)）。

　(二) 欲註明出處，但行文已有作者姓名時，按（年代：頁數）之格
　　　式加註，如：「Fine 指出……，……(1994: 11)。」。

九、縮寫、圖表名稱及註解位置

　(一) 縮寫可使用約定俗成之簡稱，但須於參考文獻中註明清楚。

　(二) 表名在表上方、表註在表下方；圖名、圖註在圖下方。圖表在
　　　文中的位置。

貳、參考文獻

　　　　曾於正文中引用之文獻，請列入參考文獻；正文未引用者，不
　　得列入。

一、編排順序

　(一) 中文資料與西文資料須分開，並按作者名稱之筆畫、字母順序
　　　編排。

　(二) 同一作者之不同文獻，以出版日期先後順序編排。

(三) 同一年之不同出版品，於年份後加上小寫英文字母以區分時間順序。

(四) 同作者之不同文獻，除第一篇須列作者全名外，其餘均以三個短破折號－－－代替作者姓名。

二、縮寫簡稱表

引用文獻之縮寫除註明於書目之中，亦可全數列於參考文獻之頁首。

三、格式

參考文獻請按以下格式列出。

(一) 期刊論文

中文期刊論文：「作者（年代）。〈論文標題〉，《期刊名》，卷，期數：頁碼。」如：

王文方（2007）。〈虛擬條件句理論述評〉，《國立臺灣大學哲學論評》，33：133-182。

西文期刊論文：Author's last name, author's first name (Year). "Title of the article." *Title of the Periodical*, series number, volume number: page numbers. 如:

LePore, Ernest., Loewer, Barry (1987). "Mind Matter." *Journal of Philosophy*, 84: 630-71.

(二) 專書論文

中文專書論文：「作者（年代）。〈論文標題〉，編者（編），《書名》，頁碼。出版地：出版者。」如：

張旺山（2001）。〈行動人與歷史世界的建造：論狄爾泰的「生命的詮釋學」〉，黃俊傑（編），《中國經典詮釋傳統（一）：通論篇》，51-75。臺北市：喜馬拉雅基金會。

西文專書論文：Author's last name, author's first name (Year). "Title of the article." Editor(s) (ed.), *Title of the book* (page numbers). Place of Publication: Publisher. 如：

Haugeland, John (2002). "Heidegger on Being a Person." Hebert Dreyfus (ed.). *Heidegger Reexamined (vol.1)*: *Dasein, Authenticity, and Death* (73-84). New York: Routledge.

(三) 專書

中文專書：「作者（年代）。《書名》。出版地：出版者。」如：

牟宗三（1983）。《心體與性體》。臺北：正中書局。

西文專書：Author's last name, author's first name (Year). *Title of the book*. Place of publication: Publisher. 如：

Allison, Henry E. (2001). *Kant's Theory of Taste: A Reading of the Critique of Aesthetic Judgment*. Cambridge: Cambridge University Press.

(四) 古籍

古籍原刻本：「作者（朝代）。《古籍標題》（版本說明），卷。」如：

劉向（清）。《列女傳》（道光 17 年振綺堂原雕，同治 13 年補刊，梁端校讀本），卷 2。

古籍複印本：「作者（朝代）。《古籍標題》（出版地：出版者，出版日期，複印版本），卷。」如：

王鳴盛（清）。《十七史商榷》（臺北：樂天，1972，影廣雅書局本），卷 12。

(五) 研討會論文

中文研討會論文：「作者（年.月）。〈論文標題〉，「研討會名稱」。地點：主辦單位。」如：

陳文團（2004.04）。〈馬克斯的正義與薪資分配制度〉,「馬克斯的正義研討會」。北京：北京清華大學。

西文研討會論文：Author's last name, author's first name (Year, Date). "Title of the article." Paper presented at the Name of the Conference, Place. 如：

Tran Van Doan (2006.11). "Harmony as the Principle of Eco-Ethics." Paper presented at The 25th International Conference on Eco-Ethics. Denmark: University of Copenhagen, Copenhagen.

(六) 博士論文

中文博士論文：「作者（年代）。《論文標題》。博士論文。發表地點（學校及科系名稱）。」如：

林雅萍（2008）。《意識、觀視與德行──梅朵論道德能動性》。博士論文。臺北：臺灣大學哲學系。

西文博士論文：Author's last name, author's first name (Year). *Title of the dissertation*. Unpublished doctoral dissertation, Name of the University, Place. 如：

Carlo Kwan (1985). *The Dimension of Time in Mythology*. Doctoral Dissertation of Philosophy. Katholieke Universiteit Leuven, Louvain.

(七) 電子出版本

中文電子出版品：「作者（年代）。〈文件標題〉,文件發表日期。電子出版品的出版資訊（出版平台名稱,主編,出版單位）。資料取得方式（取得日期,網址）。」如：

凡國棟（2009）。〈也說《凡物流形》之「月之有軍（暈）」〉,2009 年 1 月 3 日。《武漢簡帛網》,陳偉主編,武漢大學：

武漢大學簡帛研究中心。2009 年 4 月 17 日，< http://www.bsm.org.cn/show_article.php?id=941>。

西文電子出版品：Author's last name, author's first name (Year). "Title of the Document." Information of Electronic Publication. Access Information (Date of access, <URL>). 。如:

Ross, Don (1997). "Game Theory." 25 Jan. 1997. Stanford Encyclopedia of Philosophy, Edward N. Zalta(ed.), Stanford University: Center for Study of Language and Information. 1 Oct. 2002, <http://plato.standford.edu/entries/game-theory/>.

Guide for Submission to
National Taiwan University Philosophical Review

1. *National Taiwan University Philosophical Review* publishes work form any aspect of philosophy, including articles, discussions, book reviews, news, and comments on articles previously published in the journal.

2. The journal is a semiannual publication, with issues coming out each March and October. Scholars from both Taiwan and abroad are welcome to submit papers.

3. All submitted papers will be sent anonymously to two experts for evaluation. Contributors will be notified of the decision in regard to their submissions by the journal's editorial board.

4. Submitted papers can be written either in Chinese or in English. Chinese manuscripts should not exceed 25,000 words (notes included) in length, and no English manuscripts over 40 pages (A4, double-spaced) will be considered.

5. It is contributor's responsibility to revise their manuscripts according to the suggestion of the editorial board.

6. Authors of published papers will receive thirty offprint copies of their articles and three copies of the issue in which their articles appear.

7. Submitted papers must be previously unpublished. Manuscripts under consideration by any other journal are unacceptable.

8. Submissions should be complete with the following information both in Chinese and English: title, abstract (no more than 500 words), and

keywords (no more than five). Papers must follow the *National Taiwan University Philosophical Review* Style sheet.

9. Please also provide the following information on a separate sheet of paper: name, institution, and position, contact address, email address, and telephone or fax number.

10. The author of a submitted article enjoys copyright privileges, but the journal retains property rights for the article. Therefore, the author has the right to republish in a collection of his / her essays, but anyone else cannot reprint, reproduce or translate articles without acquiring permission from the journal.

11. The contributor may not infringe upon the rights of any third party through plagiarism, reproduction, or in any other fashion. The journal will not be legally responsible for such acts.

12. When submitting a paper, please send three printed copies and one electronic copy of the manuscript in Word format. The contributor should execute the copyright license agreement in duplicate originals, each agreement shall be respectively preserved for reference by the editorial board of the journal and National Taiwan University Press Center.

13. The author should agree to grant a free license of the article which were published in *National Taiwan University Philosophical Review*, to the editorial board of the journal and National Taiwan University Press Center, and other organizations which co-operated with and authorized by the editorial board of the journal or National Taiwan University Press Center to edit the article into periodical (whether in paper or in electronic form), and duplicate, translate, transmit publicly the article, and archive

the article in database, etc., and also allow users to read, download, and print, etc. The author should also agree to grant a license to National Taiwan University Press Center to publish the article in paper and electronic form for the purpose of encouraging the academic research.

14. All editorial correspondence should be addressed to the following address: Editorial Board, *National Taiwan University Philosophical Review*, Department of Philosophy, National Taiwan University, No. 1, Section 4, Roosevelt Rd., Taipei, Taiwan 10617, R. O. C.; or to the following email address: ymlai@ntu.edu.tw.