

論羅爾斯的穩定性論證

吳澤玫*

摘要

羅爾斯在前後期的正義理論裡，系統地探討並嘗試解決穩定性問題，即在當代社會存在合理多元的事實下，如何促使公民自願接受並遵守社會制度的相關規範，以維繫正義且秩序良好的社會？在《政治自由主義》裡，羅爾斯提出重疊共識的論證。他主張持不同合理全面性學說的公民都會同意「作為公平的正義」這種政治正義觀，以之規範社會主要制度。公民將基於這樣的共識，自願遵守相關規範，因而能在多元價值之中確保社會的穩定。

本文的目的在於探討重疊共識的論證能否成功解決多元社會的穩定性問題。首先說明羅爾斯關切的穩定性問題為何，其次說明他為什麼放棄《正義論》裡的穩定性論證，改採重疊共識的論證。第三，本文將分析羅爾斯如何證成重疊共識，主張這樣的共識可以解決穩定性問題。第四，本文檢視並批評重疊共識的論證。筆者將指出，這個論證是失敗的，任何採取社會正義原則層次的論證都無法成功解決穩定性問題。最後，本文將提出一種奠基於道德心理的穩定性論證之初步形式。這樣的論證取向不但相容於羅爾斯的觀點，且是理論上可以成立的。

關鍵詞：穩定性問題、羅爾斯、作為公平的正義、重疊共識、道德心理

* 吳澤玫，國立臺灣大學哲學系助理教授。

投稿：107年7月11日；修訂：107年9月15日；接受刊登：107年10月2日

On Rawls's Arguments of Stability

Tse-Mei Wu*

Abstract

John Rawls systematically explored and tried to solve the problem of stability; that is, given the fact of reasonable pluralism, how can citizens be motivated so that they will accept and comply with social institutions voluntarily to maintain a just and well-ordered society? In *Political Liberalism*, Rawls proposed the argument of overlapping consensus. He argued that citizens who affirm different reasonable comprehensive doctrines would agree to “justice as fairness” and use this political conception of justice to regulate major social institutions. Based on this consensus, citizens will voluntarily comply with relevant norms and ensure social stability.

This essay shall explore whether the argument of overlapping consensus can successfully solve the problem of stability in a pluralistic society. First, I shall explain the problem of stability with which Rawls was concerned. Second, I shall explain why Rawls gave up the argument for stability in *A Theory of Justice* and why he proposed the argument of overlapping consensus. Third, I shall analyze how Rawls argued that such a consensus can solve the problem of stability. Fourth, I shall examine and criticize the argument of overlapping consensus. By exploring how this argument fails, I shall point out how any argument appealing to principles of social justice fails to solve the stability

* Assistant Professor, Department of Philosophy, National Taiwan University.

problem successfully. Finally, I shall propose a preliminary form of the argument for stability based on moral psychology. This argument is not only compatible with Rawls's claims but is also theoretically defensible.

Keywords: the problem of stability, John Rawls, justice as fairness, overlapping consensus, moral psychology

論羅爾斯的穩定性論證*

吳澤攻

壹、前言

穩定性問題 (problem of stability) 要探討的是：如何促使公民願意接受並遵守社會制度的相關規範，以維繫正義且秩序良好的社會？羅爾斯 (John Rawls) 在前後期的正義理論裡，系統地探討並嘗試解決這個問題。¹後期他注意到當代社會存在多元的價值觀，使得穩定性問題變得更棘手。因為可能影響個人遵守意願的因素除了自身利益外，各種分歧、甚至相互對立的價值觀更將成為破壞社會穩定的深刻原因。當公民出於不同價值觀，而對社會制度持對立意見時，要如何確保他們願意接受並遵守這些制度的規範，便成為相當困難的挑戰。

穩定性問題對政治哲學理論與當代民主社會實作都有重要的研究價值。首先，羅爾斯指出，儘管在道德哲學史裡，這個問題不太受到關注，但它「卻是政治哲學的基本問題」(PL: xvii)。為什麼政治哲學家更應該關心穩定性問題？這是因為，政治哲學的理論提議涉及制度落實，而

* 本文之初稿曾於 2013 年 10 月 26 日「台灣哲學學會 2013 年學術研討會：語言、詮釋與公共理性」中發表。感謝評論人許漢教授、多位與會學者以及本文兩位匿名審查人給予豐富的修改意見，讓本文的論點能更加完善。

¹ 羅爾斯的政治哲學思想經歷重要轉變，本文以他的兩本重要著作來劃分不同時期：「前期」以 1971 年的《正義論》(*A Theory of Justice*) 為代表，以下簡稱此書為 TJ。「後期」以 1993 年的《政治自由主義》(*Political Liberalism*) 為代表，以下簡稱此書為 PL。

推動和改革制度須投注大量社會成本，所以會要求政治哲學理論必須具備可行性 (feasibility)，以使得落實該理論所投入的制度性努力會是值得的。如果一個理論能證成由它的原則所規範的社會將是穩定的，將可提昇該理論的說服力。(Hill, 1994: 338) 因此，穩定性問題在政治哲學的理論層面有重要的研究價值。

此外，當代社會面臨的一大挑戰是，政府許多重要政策提案往往引發嚴重爭議。在台灣，近年因為《都市更新條例》造成的迫遷問題和《勞基法》修正工時和例假休假規定，都是典型的例子。當持對立立場者彼此惡言相向，甚至在抗爭行動中發生流血衝突，將造成民心不安與社會秩序的混亂。對任何社會來說，維持穩定都是重要目標之一，否則可能危及社會的發展。因此，穩定性問題在民主政治實作層面也有重要的研究價值。

羅爾斯在《正義論》(*A Theory of Justice*) 的第三部分，以超過兩百頁的篇幅建構論證，嘗試解決穩定性問題。但他後來表示不滿意自己的論證，因而在《政治自由主義》(*Political Liberalism*) 裡，以「重疊共識」(overlapping consensus) 的論證取代前期論證。簡單來說，重疊共識的概念指的是，持不同價值觀的公民都會同意以「作為公平的正義」(justice as fairness) 這種政治正義觀 (political conception of justice)，來規範社會主要制度。人們將基於這樣的共識，自願遵守相關規範，因而能夠在多元價值之中確保社會穩定。

本文的目的在於探討羅爾斯的重疊共識論證能否成功解決多元社會的穩定性問題。第貳節藉由比較他和霍布斯 (Thomas Hobbes) 對穩定性的不同看法，來說明羅爾斯關注的穩定性問題具有哪些特徵。第參節簡單說明《正義論》的穩定性論證後，主要將探討羅爾斯為什麼放棄前期論證，改採重疊共識論證的原因。第肆節分析羅爾斯如何證成「作為公

平的正義」可以成為重疊共識，並訴諸這樣的共識來解決穩定性問題。第五節進一步檢視並批評重疊共識的論證。本文將指出，這個論證是失敗的。並且，任何採取社會正義原則層次的論證取向都不可能成功。最後，筆者將嘗試提出一種奠基於道德心理（moral psychology）觀點的穩定性論證之初步形式。這樣的論證取向相容於羅爾斯的觀點，且在理論上可以成立的。

貳、穩定性問題

在探討羅爾斯的穩定性論證前，首先必須釐清他對穩定性問題的設定。對羅爾斯而言，「正義」是社會制度的首要價值，但確保穩定也是建構美好社會的先決條件之一。（TJ：3、6）他後期更明確指出，政治自由主義嘗試回答的根本問題是，在存在多元價值的既定事實下，由公民組成的「正義且穩定的社會如何可能長久存續？」（PL：4）由此可以看出，社會穩定是羅爾斯非常關心並嘗試解決的重要問題。在政治哲學史裡，另一個致力探討社會穩定問題的哲學家是霍布斯。他們對這個問題的理解有三方面的差別：首先是適用的社會類型，其次是維繫穩定的力量來源，最後則涉及個人遵守社會規範的理由。這樣的 Understanding 使得他們提出不同的穩定性論證。以下從這三方面做比較，指出羅爾斯關注的穩定性問題具有哪些特徵。

一、良序社會（well-ordered society）的穩定

羅爾斯和霍布斯為穩定性問題所設定的社會類型並不相同。霍布斯是從全然缺乏行為規範的自然狀態出發，探討個人為什麼會透過社會契約來建立規範。在《利維坦》（*Leviathan*）裡，霍布斯描述在自然狀態裡，

人們出於競爭、猜疑或榮耀，為了追求個人最大利益、安全或名聲而使用武力相互侵犯，因而生活在人與人相互為敵的戰爭狀態。(Hobbes, 1998: 83-84) 為了脫離這種難以安心生活的恐怖環境，理性促使人們提出一些和平條款——自然法則 (Laws of Nature)，進而就這些法則如何規範彼此的行為達成協議，以確保所有人都可以過和平且安全的生活。

相較於霍布斯式的非理想社會，羅爾斯針對的是理想的「良序社會」。良序社會是受公共正義觀有效規範的社會，且具備兩個特徵：在這樣的社會裡，(1) 每個人都接受且知道其他人接受相同的正義原則；並且(2) 基本社會制度一般滿足了這些原則且這個事實為人們所知。(TJ: 4-5; PL: 35) 良序社會的特徵意味著，羅爾斯關心的社會穩定不只是人們遵守法律，讓社會秩序持久，還必須是一種穩定的正義 (stably just)，意即該社會能保持正義或接近正義的狀態。換言之，他要探討的是「正義的」理想社會如何能長久存續的問題。(Weithman, 2011: 44; Hill, 1994: 338) 儘管沒有任何現存社會是真正良序的，但羅爾斯認為，從理想理論的層次出發，能讓我們更清楚看出解決穩定性問題的適當方式。

二、穩定力量來自個人和社會內部

根據羅爾斯對穩定性的描述，可以瞭解他如何說明維繫社會穩定的力量來源。他說道：

社會的合作系統必須是穩定的：它的各種基本規則必須或多或少規律地被遵循，且人們自願地按照這些規則來行動；並且當發生違反的情況時，應存在穩定化的力量，以阻止進一步的違反並傾向恢復該制度措施。(TJ: 6)

公民自願地行動，以給予彼此長期的正義。穩定性是由在正義的制度下獲得的適當種類的充分動機所確保。(PL：142-143)

對羅爾斯而言，社會的穩定來自個人擁有適當的動機，自願遵守規範。並且，穩定的社會並不意味著沒有人會違反規範，而是即使發生這種情況，也存在有效的矯正力量，使社會回復到原來的良好狀態。就維繫社會穩定的力量而言，羅爾斯的觀點不同於霍布斯。

霍布斯認為，儘管自然狀態裡的人們同意以自然法則來規範彼此的行為，但這些法則本身不足以確保所有人都會履行協議、遵守相關規範，於是可能產生搭便車（free-rider）的問題。搭便車指的是，當社會成員普遍遵守規範，某個人不遵守卻仍享受合作帶來的好處，由此獲取最大個人利益。這對社會穩定將構成嚴重威脅，因為一旦其他人發現有搭便車者，他們將不願繼續遵守規範，於是可能使社會陷入失序的狀態。霍布斯主張，人們必須把自己的權力和力量授與一個共同權力，讓它透過懲罰的威嚇來把所有人的意志組織為一個人格（person）。霍布斯把承擔這個人格的對象稱為主權者（sovereign），由此建立的國家就是他所謂偉大的「利維坦」。(Hobbes, 1998：114) 允許主權者嚴厲懲罰違規者，可以確保所有人都確實遵守社會規範。

羅爾斯不訴諸外在懲罰來避免搭便車的問題，他主張在良序社會裡，消除少數違規情事的力量來自社會內部。任何系統的穩定性取決於可幫助它回復均衡狀態的內部力量強度。(TJ：457) 良序社會作為一種合作系統，它的均衡狀態指的是各種主要制度符合正義且可有效落實，讓所有人都享有正義的對待。在一個穩定的社會裡，當某些人違反規範或因其他因素導致社會制度偏離正義時，在其內部將出現矯正力量，使社會整體秩序回復到原來的正義狀態。這樣的內部力量越強，社會就越

穩定。相較於霍布斯訴諸主權者的外在制裁，羅爾斯以正義感（sense of justice）這種道德情感，來說明內在於個人的穩定力量。²正義感指的是「一種運用並按照正義原則來行動，且因此是從正義的觀點來行動的有效欲望」。（TJ：567）良序社會的另一個特徵是，該社會的公民擁有正常有效的正義感，這種正義地行動的內在欲望不但構成個人遵守社會規範的有效動機，且能在出現違規情事時發揮矯正力量，故得以確保社會的長期穩定。（PL：35）

三、出於正當理由的穩定（stable for the right reasons）

根據霍布斯的觀點，個人為了脫離難以安心生活的自然狀態而同意自然法則，且因為害怕被主權者懲罰，所以遵守根據自然法則制訂的社會規範。對羅爾斯而言，基於這樣的理由而遵守規範只能達成奠基於「暫訂協議」（*modus vivendi*）的短期穩定，無法長久維持。暫訂協議的典型用法出現在兩個國家的目標和利益發生衝突時，雙方出於自身利益的考量，根據彼此政治權力的現存平衡點而達成協議。儘管遵守協議對各自都有好處，不過一旦權力的平衡發生變化，取得較大優勢的一方將不再願意遵守協議。（PL：147）衛科（Aaron Weddle）指出，在社會穩定的問題裡，可以把上述的「國家」代換為持對立價值觀的不同學說。如果這些學說的支持者僅是出於自身利益及現有政治權力的考量而達成協議，那麼這不是真正的穩定。（2017：219）因為一旦權力結構改變，獲得較大權力的學說支持者將不會繼續遵守規範，反而會嘗試改變社會制

² 崩斯諾夫（Larry Krasnoff）認為，根據理想理論和非理想理論的區別，可以說明羅爾斯和霍布斯的穩定性論證的差別。對羅爾斯來說，對違規者施以強制性懲罰的社會系統屬於非理想理論，他認為穩定性問題可以且應該由理想理論來處理。換言之，儘管個人可以非理想地出於對懲罰的恐懼而遵守規範，也應該能理想地基於正義的行動符合理性考量而遵守。（2015：805-806）

度，使之符合自身價值觀。若此，則社會將陷入權力競逐導致的不穩定狀態。

前文曾提到，對羅爾斯而言，維繫社會穩定的力量來自個人擁有適當的動機。這樣的適當性跟遵守規範的理由有關。他主張良序社會欲追求的是「出於正當理由的穩定」。(1996: xxxix) 這是說，人們是基於「道德的」理由自願遵守規範，而不是因為自己的政治權力太弱或害怕遭受懲罰，所以才被迫遵守。本文稍後將指出，在重疊共識的論證裡，羅爾斯主張「作為公平的正義」能夠確保這種出於正當理由的穩定。

參、前期穩定性論證及其錯誤

根據以上所述，我們可以瞭解到，羅爾斯關注的穩定性問題之適用對象是正義的良序社會，維繫穩定的力量來自社會成員擁有的內在動機，目標則是尋找出於道德理由所達成的社會穩定。他在《正義論》裡提出包含兩個部分的穩定性論證：正義感的論證和契合性論證（congruence argument）。³由於本文關注的焦點在於後期的重疊共識論證，所以底下僅簡單說明這兩部分論證的要點，為的是探討羅爾斯為什麼放棄前期的論證，而以重疊共識的論證加以取代。

一、正義感的論證

在穩定性論證的第一部分，羅爾斯嘗試證明，在受「作為公平的正義」的兩個正義原則有效規範的良序社會裡，人們能夠發展出強烈的正義感，這將使得他們擁有適當的道德動機，願意按照正義原則的要求行動。羅爾斯首先提出道德心理主張，來說明正義感的發展歷程。

³ 「契合性論證」一詞出自傅理曼（Samuel Freeman）。（2003：277）

人類的道德發展經歷三個階段：權威的道德 (morality of authority)、結社的道德 (morality of association) 和原則的道德 (morality of principle)。每個階段相應於不同共享制度，依序是家庭、結社團體，最後則參與在社會的各種制度裡。對羅爾斯來說，人的道德心理會受到成長與生活脈絡的道德性影響。(Baldwin, 2008: 248) 在一個正義的良序社會裡，個人在孩童時期感受到父母的無條件關懷，參與結社團體時看到其他成員善盡職責，將感受到自己由此獲得的利益，形成對父母的愛以及對結社成員的友誼與信任。於是傾向接受父母給出的規則，遵守結社訂立的規範。隨著道德能力逐漸成熟，個人在持續應用和解釋正義標準的過程裡，瞭解規範社會的正義原則。當他在社會生活裡體認到正義的社會制度讓自己及關心的人都獲益，將發展出對正義原則的認同，以及按這些原則行動的強烈欲望 (即正義感)。(TJ: 462-479) 這三個階段的共通基本概念是相互性 (reciprocity)，即以善意回應善意的傾向。(TJ: 494) 這是說，個人認知到父母、結社成員和社會制度為了自己的利益而行動的明顯意圖，所以他也會關心他人的福祉、遵守相關規範作為回報，違反時將產生嚴重罪惡感。我們可以說，個人在各階段知覺到自己的福祉如何受到他人和社會制度的影響，因而逐漸形成正義感。

正義感會以兩種方式展現：首先，它引導個人接受正義的制度，希望盡自己的本分來維繫這些制度，並根據共同接受的正義原則來裁決歧見。第二，正義感引發人們願意為建立正義制度而努力，以及依正義的要求來改革現有制度的意願。(TJ: 474) 就此而言，擁有正義感使得人們不會希望以違反正義原則的方式來促進自身利益，否則就會受到罪惡感的譴責。並且，當發現有人違反正義原則時，正義感將促使個人採取行動，試圖規勸或矯正錯誤行為。(TJ: 497-498)

羅爾斯進一步指出，相較於其他正義觀，「作為公平的正義」的兩個正義原則能讓人發展出較為強烈的正義感，所以是比較穩定的正義觀。這是因為，第一正義原則（平等自由權原則）要求讓所有人都享有各種基本權利，確保個人不會因為社會整體利益而被犧牲；第二正義原則包含的機會公平平等原則和差異原則（difference principle）排除來自社會和自然的偶然因素對分配的影響，這些都展現出對個人的無條件關懷。此外，兩個正義原則及其道德觀相當清晰，不但易於瞭解也能得到公共認可。因此，生活在受兩個正義原則有效規範的社會，能促使人們發展出較強烈的正義感。（TJ：499-501）由於良序社會存在這種有效的內在穩定力量，所以是比較穩定的社會。

二、契合性論證

前期穩定性論證的第二部分是契合性論證，羅爾斯嘗試證明正義原則的要求符合個人的「好」（the good）。⁴對此可能有的疑惑是：根據良序社會的定義，在這種理想社會裡，每個人都接受相同的正義原則。並且，根據正義感的論證，在受兩個正義原則規範的社會裡成長的人，將發展出強烈的正義感，那麼為什麼還需要論證「正義」和「好」的契合性？這是因為，即使良序社會成員擁有正義感，但他們不見得在所有情況下都會堅持履行正義感，仍可能違反規範。更重要的是，假如正義原則的要求對個人不利，或者跟他想要實現的其他重要目標發生衝突，便難以期待人們始終會堅持依正義原則的要求來行動。因此，羅爾斯還要

⁴ “the good”一詞指稱非道德性的價值，“the right”則指稱跟個人行為和社會正義有關的道德性價值。為了避免混淆道德與非道德性的價值，故本文不把“the good”譯為「價值」或「善」，而是譯為「好」。

進一步證明，保持自己的正義感、堅持遵守正義原則，對個人來說是好的，這樣的契合性可以形成遵守規範的更充分動機。

羅爾斯指出，個人的「好」是由他的理性生活計畫（rational plan of life）所決定，指的是當個人知道相關事實，仔細考慮各種選項的可能結果後，根據理性選擇原則選出的長期計畫。（TJ：408-409）⁵追求並實現理性生活計畫所鼓勵的目標，對個人來說是好的。於是，若要證明「正義」和「好」的契合性，就必須證明「對良序社會裡的那些人來說，肯定他們的正義感作為其生活計畫的規範因素是……理性的」（TJ：567）。如此便可期待儘管人們有不同的理性生活計畫，但會一致地肯定並遵守兩個正義原則。

羅爾斯提出三個理由來論證，把正義感視為理性生活計畫的規範因素，堅持遵守正義原則，對個人是好的。首先是訴諸違反規範讓個人付出的心理代價。在良序社會裡，正義原則公開為人們所知，它於是描述了社會成員共享的道德信念。就此而言，那些假裝自己擁有正義感（但實際不然）的搭便車者必須採取一些預防措施，來維持欺騙和偽善的狀態，且必須接受由此造成自己喪失行動的自發性（spontaneity）和自然性的代價。（TJ：570）此外，所有人都希望自己愛的人能得到公平對待，但基於社會制度和個人的密切連結，那些搭便車者無法選擇誰會因為自己的違規行為受到傷害，對此形成的擔憂也將成為重大的心理代價。相反地，如果個人保持正義感，便可以自然且簡單的方式，避免付出這些代價。（TJ：571）因此，肯定自己的正義感作為理性生活計畫的規範因素，對個人來說是好的。

⁵ 羅爾斯以三階段的方式來定義「好」。（TJ：399-403、408-415）由於這部分的主張並非本文關注的焦點，在此僅說明跟契合性論證最相關的第三階段定義。

其次，無論個人選擇的理性生活計畫為何，參與良序社會的生活對每個人都有重要價值。這是因為，人類具有社會性（sociability），他們會參與各種結社團體——羅爾斯稱之為「社會統合體」（social union）。這樣的社會本性不僅意味著藉由合作可以讓個人滿足自己的需要和利益，也展現出人們擁有共享的終極目的——成功施行各種正義制度，肯定這些制度本身是好的。因為人們瞭解到，唯有當各個結社受到正義原則的良好規範，他們才能在結社活動裡實現自己的潛能和目標。（TJ：522、527-529）正義原則的有效規範把各個結社及其活動統合起來，讓人們在合作中提昇福祉，並享受更豐富且多樣的集體活動。個人於是體認到，如果要充分共享這種有價值的社會生活，必須認可相同正義原則的規範。（TJ：571）由此可以推論，肯定自己的正義感，堅持按正義原則的要求來行動，對個人來說是好的。

最後，透過「作為公平的正義」的康德式詮釋（Kantian interpretation），可以支持遵守兩個正義原則跟個人的「好」相互契合。在這裡，康德（Immanuel Kant）的自律（autonomy）概念發揮了關鍵的論證作用。對康德而言，人類做為自由、平等的理性存有者，其獨特之處就在於能夠擺脫各種偏好、欲望和利益的考量，選擇可普遍化的行為準則作為道德法則。由於道德法則的約束力來自人的理性意志，所以遵守道德法則就是遵守自己制定的法則，因而是自律的行動。羅爾斯主張，遵守「作為公平的正義」的正義原則讓人得以展現自律的行動。這是因為，在原初立場（original position）裡，無知之幕（veil of ignorance）排除立約者對其社會地位、自然稟賦和個人利益的認識，所以他們對正義原則的選擇不會受到這些偶然因素的影響。於是，依據最後選出的正義原則來行動，就是自律的行動。（TJ：251-252）根據康德式詮釋，正義地行動的欲望（展現正義感），跟表達自己做為自由平等道德人本性的欲望，是相同的

欲望。因為它們都是按照在原初立場中將被選擇的原則（兩個正義原則）來行動的傾向。（TJ：572）換一種方式來說，只有依據正義原則來行動，才能實現人類的本性。羅爾斯由此得出以下結論：「因此，為了實現我們的本性，我們別無選擇，只能計畫去保存我們的正義感，用來規範我們的其他目標。」（TJ：574）必須澄清的是，對羅爾斯來說，保持正義感並依據兩個正義原則自律地行動之所以對個人是好的，不是建立在工具價值的考量（即不是因為這麼做可以為個人帶來特殊利益），而是因為它實現了人類的最高價值——表達自己做為自由平等的理性存有者的本性。這是一種內在價值，對所有人來說都是好的。

三、合理多元事實（the fact of reasonable pluralism）的挑戰

羅爾斯後來表示不滿意自己的穩定性論證，理由是：「《正義論》第三部分對穩定性的說明跟整體觀點不一致……。」（PL：xvii-xviii）他對不一致的問題做了這樣的解釋：

我內心以為的嚴重問題涉及出現在《正義論》裡的良序社會概念是不切實際的（unrealistic）。跟作為公平的正義有關的良序社會，它的一個基本特徵是：它的所有公民是在我現在稱為全面性哲學學說（comprehensive philosophical doctrine）的基礎上，認可作為公平的正義這個觀念。（PL：xviii）

羅爾斯指出，當一個觀念包含人生的價值、個人德行與性格的理想，而對整體生活的行為提供指導時，它就是全面性的觀念。（PL：175）在《正義論》對良序社會的相關說明裡，涉及全面性觀念的是契合性論證的第三個理由。前文提到，根據「作為公平的正義」的康德式詮釋，唯有發揮正義感，依照兩個正義原則自律地行動，才能實現理性

存有者的本性。由於康德對自律和人類最高價值的主張欲規範生活裡的所有行為，所以他的理論是一種全面性學說。契合性論證若要成立，則所有社會成員都必須接受康德的道德理論。然而，當代社會存在的「合理多元事實」使得這種觀點成為不切實際的。(Freeman, 2003: 304)

合理多元事實指的是，有許多不同的宗教、道德和哲學的全面性學說，這些學說各有其特殊的理論主張和價值觀。並且，其中許多都是「合理的」學說。由於合理多元事實是民主社會公共文化的一個主要特徵(PL: 36)，這將導致契合性論證的失敗。因為用來支持這個論證的重要理據——所有人都接受康德的全面性學說——無法成立。我們可以說，羅爾斯認為前期論證的錯誤在於：它要求公民根據康德的全面性哲學學說，肯定「作為公平的正義」跟個人的「好」相互契合。但適用於民主社會的正義理論必須承認全面性學說的合理多元事實，因此，自己對穩定性的說明跟理論的整體觀點不一致。這個理論內部的問題減弱了「作為公平的正義」的穩定性，於是連帶破壞羅爾斯試圖以穩定的正義觀來達成社會穩定的理論效力。

儘管不滿意《正義論》裡的穩定性論證，但羅爾斯依然相信「作為公平的正義」是穩定的正義觀，可以確保社會穩定。為了證明這一點，他在後期提出新的穩定性論證——重疊共識，⁶這是接下來要深入探討的課題。

⁶ 威特曼 (Paul J. Weithman) 認為重疊共識是創新的概念，不是對前期正義論的澄清。(2011: 37)

肆、重疊共識的論證

羅爾斯後期做出理論修正，嘗試化解其整體觀點跟穩定性說明不一致的問題。為了把自己的正義觀跟康德的全面性學說區隔開來，他明確地把「作為公平的正義」定位成「政治的」正義觀。這樣的正義觀有三個特點：首先，它的適用對象是當代採憲政民主政體的社會基本結構，而不適用個人行為或國際關係。其次，它是獨立自恃的（freestanding）觀點。這是說，它的內容不涉及全面性學說主張，無意探討人生價值、個人德行與性格等理想，本身不是全面性學說。在證成其觀點時也不訴諸任何全面性學說及相關的形上學、知識論主張。第三，它的內容來自憲政民主社會的三個共享觀念，包括：把社會看成公民世代公平合作的系統、公民做為自由平等人的「政治性人觀」（political conception of the person）、把良序社會視為受一種政治正義觀有效規範的社會。這些觀念不涉及全面性學說，是所有社會成員都能共同接受的。（PL：11-15）

根據政治正義觀的這些特性，我們可以看到，羅爾斯關心的穩定性問題在適用社會類型方面做了限縮：他想要解決當代憲政民主社會的穩定性問題，而不是要提出適用所有時代和各類政體的穩定性論證。此外，我們也可以瞭解到，政治正義觀之所以必須是獨立自恃的觀點，是因為當代社會存在合理多元的事實，任何全面性學說都不可能得到普遍支持。如果要讓抱持不同宗教信仰、道德理想或特殊價值觀的人們共同接受政治正義觀的規範，那麼它的證成方式和內容不能依賴任何全面性學說，而必須訴諸共享的觀念。

羅爾斯後期的穩定性論證嘗試證明，一旦「作為公平的正義」被呈現為政治的正義觀，它便能贏得所有社會成員的認可（無論他們抱持哪

種合理全面性學說)，羅爾斯稱之為「重疊共識」。以下進一步說明這個論證。

一、什麼是重疊共識？

必須先澄清的是，羅爾斯提出重疊共識的概念，要論證的是他的政治正義觀的穩定性，而非該正義觀的內容。「作為公平的正義」在推論內容時是從三個共享觀念出發，先透過原初立場的理論設計，選出最具正當性的原則——兩個正義原則；而後再以「重疊共識」的概念來證成，這樣的政治正義觀能獲得所有社會成員的認可，因而將可確保社會穩定。兩階段的證成有密切關連：第一階段為政治正義觀確立的獨立自恃特性，促成第二階段達成重疊共識的可能。

其次，羅爾斯要尋求的是「合理的」全面性學說之間的共識，而非主張現存的所有全面性學說都能認可他的政治正義觀。(PL:144)也就是說，重疊共識的概念不考慮不合理的學說。羅爾斯指出，合理的全面性學說有三個主要特徵：(1) 理論理性 (theoretical reason) 的運用：該學說以一致且融貫的方式來說明各種跟宗教、哲學與道德面向有關的價值，並且表達出可理解的世界觀。(2) 實踐理性 (practical reason) 的運用：在挑選重要的價值、平衡或排除重要價值之間的可能衝突時，具備應用上的合理性。(3) 學說主張的穩健發展：它通常屬於或源自一種思想和學說的傳統，即使觀點有所改變，也是基於該學說認為良好且充分的理由緩慢地進展。羅爾斯表示，他刻意以較鬆散的方式來定義合理全面性學說，避免在缺乏有力根據的情況下，武斷地排除某些學說。(PL:59-60) 當代社會有許多全面性學說都符合上述條件，它們各自擁有相當數量的支持者，卻有著分歧、甚至是對立的觀點。

重疊共識的論證是從這樣的基本想法出發：在良序社會裡，儘管公民抱持不同的合理全面性學說，因而有不同的特殊價值觀，但他們能夠根據其學說內部提供的理由，同意根據「作為公平的正義」這種政治正義觀來規範社會主要制度。羅爾斯後期把公共和非公共領域明確劃分開來，政治正義觀屬於公共領域，全面性學說屬於非公共領域。他認為合理全面性學說對人類生活的價值、家庭與結社關係的理想所抱持的分歧主張，可以被保留在非公共領域。在公共的政治領域裡，各學說能對「作為公平的正義」取得共識。

二、典型例子 (model case)

為了說明合理全面性學說如何能從學說內部給出支持理由，對「作為公平的正義」形成重疊共識，羅爾斯以四種學說作為「典型例子」，包括：全面性宗教學說、康德的道德學說、古典效益主義 (classical utilitarianism) 和多元主義的觀點 (pluralist view)。⁷以下分別說明羅爾斯如何論證這些學說的支持者會同意他的政治正義觀。

(一) 全面性宗教學說

羅爾斯指出，由於合理的宗教學說肯定信仰自由，所以持這些宗教觀點的人將會支持容忍原則和憲政體制保障的各種基本自由。(PL: 145) 蒯斯諾夫 (Larry Krasnoff) 對此做進一步的解釋：對宗教信仰者來說，真正的價值是宗教價值，但以強迫的方式無法傳達這種重要價值。儘管

⁷ 《政治自由主義》裡二度談到重疊共識的典型例子，其中列舉的學說略有不同。第七講第3節列舉的三種觀點包括：全面性宗教學說、康德和彌爾 (John Stuart Mill) 的全面性自由主義道德學說、多元主義的觀點。(PL: 145-146) 然而，當第8節再次談到典型例子時，則指明四種觀點，包括：康德的道德學說、以邊沁 (Jeremy Bentham) 和西季威克 (Henry Sidgwick) 為代表的古典效益主義學說、多元主義的觀點、全面性宗教學說。(PL: 169-170) 本文針對後者包含的四種學說進行說明和分析。

政治價值並非終極價值，但正是基於這個理由，他們會認為政治必須擱置終極價值來自哪裡的問題，為人們理解宗教價值保留充裕的空間。(1998:288) 因此，合理宗教學說的支持者會肯定一種在內容和證成方面都獨立於宗教價值的政治正義觀。由於「作為公平的正義」是獨立自恃的正義觀，且提倡容忍原則、充分保障各種基本自由(包括信仰自由)，所以他們會同意以這種政治正義觀來規範社會主要制度。

(二) 康德的道德學說

儘管羅爾斯後期明確地把「作為公平的正義」定位成政治的正義觀，但他認為康德的道德學說跟他的正義觀之間仍有演繹的(deductive)關係，意即從康德學說的內部主張便可推論出「作為公平的正義」。(PL:169) 這樣的演繹關係不難理解，前文指出，根據康德對人類本性和自律的觀點，透過原初立場的理論設計就可推論出羅爾斯的兩個正義原則。因此，支持康德學說的人會同意以這種政治正義觀來規範社會主要制度。

(三) 古典效益主義

效益主義的基本主張是：道德評價的對象包括個人(或群體)的行為和社會制度，行為對錯或制度正義與否的判準取決於其結果對整體福祉的影響。在羅爾斯的理解中，古典效益主義把福祉定義為「理性欲望的滿足」。(TJ:25) 於是，它將根據社會制度能否讓所有人的理性欲望獲得最大淨值的滿足，來判斷該制度是否符合正義。(TJ:22)

羅爾斯認為，效益主義者一旦考量到「我們對社會制度一般而言有限的知識以及對持續發展的環境的知識……法律與制度的規則之複雜限制，以及公共理性的指導方針必須具備簡單性」等因素，便會同意「作為公平的正義」。(PL:170) 羅爾斯對此沒有做更多說明，但我們可以

在《正義論》裡找到相關論據。對羅爾斯來說，古典效益主義以「理性欲望滿足淨值的極大化」來判斷社會制度，這樣的標準相當模糊且不明確。因為要計算某項政策對欲望滿足總量的影響，必須對不同人的福祉函數做出估計，並建立人際比較的標準。但我們很難對此提出客觀的標準。如果人們對福祉計算結果有不同意見，效益主義者也無法提出更基礎的原則來說明，該如何對這種歧見做出合理裁決。相對地，以「作為公平的正義」的差異原則為例，它以社會基本價值物（social primary goods）作為衡量社會最不利者的指標。⁸我們只要看欲檢視的社會制度是否能讓這群人獲得更多基本價值物，就可以確定該制度是否符合正義。（TJ：320-321）此外，差異原則也避免了做人際比較的困難。因為只要能辨認出社會最不利者，就不必計算這些人跟其他人之間的差距有多少。（TJ：91）

羅爾斯主張，古典效益主義和「作為公平的正義」之間具有近似（approximation）的關係。這是說，效益主義者會認為，「作為公平的正義」是最接近效益原則要求的可行選項。（PL：170）他們將基於這種政治正義觀在應用上較為簡單的考量，同意以之作為政治領域的規範原則。

（四）多元主義觀

多元主義觀是一種部分全面性的學說，意即它不主張有任何一般且具支配性的單一原則或原則的組合，可用來規範整體生活。不同價值領域可適用不同原則。（Mittiga，2017：412）人類肯定許多價值，除了政治正義觀闡述的各種政治價值外，還有龐大的非政治價值。這些價值分屬不同領域，人們不需要也不可能把各類價值組織成一個融貫整體。羅

⁸ 社會基本價值物指的是「每個理性的人都被假設想要的事物。無論一個人的理性生活計畫是什麼，這些價值物通常是有用的。」社會基本價值物包括：權利與自由、權力與機會、所得與財富、自尊。（TJ：62）

爾斯認為，儘管在多元主義的學說裡包含政治價值和非政治價值，但其支持者會認為，在使得民主成為可能的合理有利條件下，政治價值通常優先於與之衝突的其他價值。(PL：145-146) 為了讓人們可以自由追求各種價值，多元主義的支持者會肯定良心自由、政治自由相對於其他價值的重要性。由於「作為公平的正義」保障人們享有這些基本權利和自由，並賦予特殊的優先性，使得多元主義論者能根據公平的條件，追求自己認為有價值的生活方式，實現其非政治價值。(PL：155) 因此，他們會同意以這種政治正義觀來規範社會主要制度。

三、奠基於重疊共識的社會穩定

羅爾斯透過四種學說的典型例子來論證，儘管社會裡存在多樣且持對立主張的合理全面性學說，但它們會認可相同的政治正義觀。利斯特 (Andrew Lister) 指出，我們可以把重疊共識的論證看成另一種形式的契合性論證，它要達成的是「政治的和全面性的正義之間的高程度契合」。(2015：138) 政治的正義指的是「作為公平的正義」這種正義觀，而全面性的正義指的則是各種合理全面性學說對規範公共領域事務的正義觀所採取的觀點。在後期理論裡，羅爾斯不但看到個人理性生活計畫呈現出不同「好」的觀念，更在當代社會的合理多元事實裡瞭解到，這些價值觀的差異來自不同全面性學說。因此，唯有透過能獲得普遍認可的政治正義觀，才能確保多元社會的穩定。

根據四個典型例子的說明，我們可以看到羅爾斯後期穩定性論證的一個重要特點：它不再訴諸單一理由（即康德全面性學說的觀點），去論證個人為什麼願意接受「作為公平的正義」這種政治正義觀。人們可根據不同理由（例如宗教學說教義、古典效益主義或多元主義提供的觀點）去決定，為什麼這種正義觀及其涉及的政治價值被賦予如此重要的地

位。由於支持理由來自個人深刻持守的全面性學說，這將使得人們更願意發自內心，遵守根據該正義觀制訂的社會規範。

然而，重疊共識的論證是否意味著羅爾斯是以政治正義觀的「可行性」——它能夠在應用層面被不同合理全面性學說的支持者接受，來決定該正義觀的「可欲性」(desirability)——即它能夠獲得規範層面的公共證成？哈伯瑪斯(Jürgen Habermas)批評，為了解決社會穩定的問題，羅爾斯把證成重疊共識的判準建立在各種全面性學說的實際接受(actual acceptance)之上，這將使得重疊共識變成促進社會穩定的工具。意即人們實際上都同意「作為公平的正義」這種正義觀，才能達成社會穩定。但我們必須先確認一個正義觀是否具備獲得證成的可接受性(justified acceptability)，而後才探討它能否得到普遍的實際接受(形成重疊共識)。羅爾斯顛倒了論證順序，這將使得他的正義觀在錯誤的意義下成為「政治的」，僅能達成暫訂協議。(1995: 119-122) 柯亨(Gerald Allan Cohen)指出，如果把穩定性理解成正義原則施行後得以持久的傾向，那麼羅爾斯把穩定性視為原則的證成條件將是荒謬的。因為「正義」和「穩定」是不同的要求。儘管人們都希望尋求能持久施行的規範原則，但不能錯誤地認為符合正義的規範原則在概念上必然蘊含穩定性。正義原則必須把穩定性納入考量，但穩定性並非用以證成該原則的條件。相反地，我們必須先判斷一個原則是否正確闡述了正義的意涵，接下來才判斷其可行性。(2008: 327-330) 周保松也提出類似批評，他認為羅爾斯把不應屬於正當性範圍的穩定性考量，視為證成正義原則的必要條件，犯了邏輯上的範疇謬誤。(2013: 163-167)⁹

⁹ 周保松認為，如果把穩定性理解為追求長久的社會秩序，會讓羅爾斯的理論面臨極大困難。他提出另一種對穩定性的詮釋，清楚區分「道德穩定性」和「社會穩定性」，主張羅爾斯首要關心的是透過實現「正義感的優先性」所達成的道德穩定性。支持正義感優先的理由是道德理由，在概念上獨立且先於支持社會穩定的理由，而合理的正義觀必須滿足道德穩定性這

對於哈伯瑪斯的批評，羅爾斯回應指出，「作為公平的正義」證成規範原則的方法不同於傳統道德理論，他的政治正義觀不涉入跟道德真理有關的觀點。用來證成原則「合理性」的整體判準，是公民擁有的「一般且廣泛的反思均衡」(general and wide reflective equilibrium)。因為它是公民在仔細考量各種正義觀選項以及支持這些正義觀的論據效力後達成的，所以是廣泛的。因為公民瞭解到，所有人的深思熟慮判斷 (considered judgments) 都認可相同的公共正義觀，所以它是一般的。(1995: 141) 在這種反思均衡裡，個人的深思熟慮判斷和原初立場的設計都可能做修正，當兩者達到融貫狀態時，正義原則就被證成是合理的。公民透過這種反思均衡，認可一種政治正義觀符合其抱持的合理全面性學說時，形成的是合理的重疊共識 (reasonable overlapping consensus)。這樣的共識並非像哈伯瑪斯所說，取決於不同學說的實際接受——這是日常政治 (everyday politics) 的共識概念，政治家經常根據現有的利益和要求，找尋能獲得所有人或多數人支持的政策。(1995: 144-145) 相對地，一般且廣泛的反思均衡涉及對正義觀和相關理據的深思熟慮判斷，所以具有規範性質，由此達成的合理重疊共識具備獲得證成的可接受性。

根據合理的重疊共識概念，我們也可以瞭解到，羅爾斯並非像柯亨所說，只是把穩定性理解成正義原則施行後得以持久的傾向，並以之作為原則的證成條件。用來證成原則合理性的判準是具規範性質的反思均衡。前文曾提到，羅爾斯的理論分兩階段來證成正義原則，儘管第二階

個證成條件。一旦接受這種詮釋，就可以把討論的重點從正義原則在應用層面的可行性問題，轉向證成層面的可欲性問題。(2013: 167-173) 筆者不同意這種對穩定性的詮釋，因為如同第貳節指出的，根據羅爾斯前後期理論對穩定性以及政治自由主義欲解決的根本問題之描述，他主要關心的是社會穩定性的問題。重點在於，達成社會穩定的方法並非如同哈伯瑪斯或周保松所說，僅著重正義原則得到實際接受的應用可行性，本文接下來將澄清這個重點並回應批評。限於篇幅的考量，在此無法對周保松的詮釋做更多討論。

段的重疊共識論證要回答的是社會穩定問題，但這不表示這樣的共識缺乏規範層面的正當性。羅爾斯指出，重疊共識的對象（「作為公平的正義」）本身是一種道德觀念，並且，從前述的典型例子可以看到，公民是根據其全面性學說所提供的「道德」理由（例如，信仰自由、自律的價值、自由追求有價值的生活方式），而認可這種政治正義觀。（PL：147）因此，由合理重疊共識達成的是「出於正當理由的穩定」，以之為正義原則提供公共證成，並沒有犯概念範疇上的謬誤。

綜合以上所述，我們可以說，重疊共識的論證並不意味著羅爾斯關心的只是政治正義觀的「可行性」（雖然這個考量也很重要），或者是以一個正義觀能否得到普遍的實際接受來證成該正義觀。並且，從重疊共識的規範性特徵（共識的對象本身是道德觀念，且奠基在道德的理由）可以瞭解到，人們不是基於各學說取得的政治權力現狀或團體利益，無奈做出的妥協，所以這種共識不是暫訂協議。於是，重疊共識論證除了嘗試因應合理多元事實，也可望消除因為全面性學說之間政治權力的變動，對社會穩定構成的威脅。由於人們對「作為公平的正義」的認可是出於道德理由的道德共識，所以即使某個學說取得較大政治權力，其支持者仍會持續遵守這種政治正義觀。

假設「作為公平的正義」確實能在多元的全面性學說與特殊價值觀之中，讓所有人對之形成合理共識，那麼它將可確保良序社會維持長久的穩定狀態。因為當個人發現某個符合該正義觀的社會制度跟自己抱持的全面性學說主張發生衝突時，他將會肯定正義觀及其政治價值具優先性，因而願意適度修正自己的價值觀或欲追求的目標，依照社會制度的要求來行動。重疊共識的論證便是透過這樣的方式，證明「作為公平的正義」可以維繫多元社會的穩定。

伍、多元社會有「重疊共識」？

當代社會的合理多元事實讓羅爾斯放棄了前期的穩定性論證，並對其正義觀的整體特性做出政治的轉向。如果要達成出於正當理由的社會穩定，不能如同霍布斯一樣，訴諸政府強制權力的壓制性使用，必須訴諸所有合理全面性學說都能認可的政治正義觀，以這樣的共識來確立遵守社會規範的道德動機。問題是，重疊共識的論證是否成功呢？

希爾（Thomas E. Hill, Jr.）指出，把「作為公平的正義」呈現為一種政治正義觀而非全面性學說，是解決穩定性問題的一個正確方向。這是因為，唯有當規範社會制度的正義觀能相容於公民抱持的全面性學說時，他們才可能接受該正義觀。而把「作為公平的正義」定位成政治的正義觀是使得二者相容的關鍵要素。因為政治正義觀不宣稱自己是真理，所以不會跟涉及真理的全面性學說相矛盾。此外，因為它是從民主社會公共政治文化的共享觀念得出內容，所以不會要求人們必須先接受特定的全面性學說，才能同意該正義觀。於是，這樣的正義觀比較可能獲得廣泛接受。（1994：340-341）就此而言，重疊共識的論證比前期的契合性論證更適合存在合理多元事實的當代社會。

然而，筆者認為，重疊共識的論證仍無法成功解決多元社會的穩定性問題。這是因為，這個論證的核心主張——「作為公平的正義」可以成為各種合理全面性學說的共識焦點——無法成立。以下提出三個理由來支持這個論點。

一、失敗的典型例子

羅爾斯主張合理的全面性宗教學說、康德的道德學說、古典效益主義和多元主義觀，都可以基於自身的理論立場同意「作為公平的正義」。要分析他對這些典型例子的論證是否成立，必須探討「作為公平的正義」的內容能否獲得這些學說的認可。政治正義觀的內容包含針對社會基本結構提出的實質正義原則，以及推論重要政治問題的探索指導方針。(PL：224)「作為公平的正義」的實質正義原則就是羅爾斯的兩個正義原則：平等自由權原則、機會公平平等原則和差異原則。由於羅爾斯主張重疊共識的廣度可以涵蓋政治正義觀的各種原則和價值(PL：149)，於是我們要問：典型例子列舉的四種學說能否根據內部的道德理由，對兩個正義原則形成合理的重疊共識？

首先，筆者同意羅爾斯所說，康德道德學說的支持者能對兩個正義原則形成共識。事實上，羅爾斯前後期對兩個正義原則的證成都受到康德理論的重要影響。差別在於後期僅是在政治的意義下，採用康德的人觀和自律概念。由於康德學說的支持者肯定所有人都是自由、平等的理性存有者，且把自律視為整體生活的最重要價值，所以他們會同意兩個正義原則是展現自律行動、實現人類本性的最適當原則。

不過，多元主義論者是否會接受兩個正義原則，是值得懷疑的。羅爾斯這裡談的是涉及形上學有關價值本質主張的「價值多元主義」(value pluralism)，而非僅是對多元價值的事實性描述。¹⁰價值多元主義對道德

¹⁰ 這裡採用的是林奇 (Sterling Lynch) 對兩種多元主義觀的區分。「價值多元主義」是以有關道德宇宙的價值本質主張，來解釋人們為什麼會有不同價值觀。這類學說的代表人物是克羅德 (George Crowder)、蓋爾斯敦 (William A. Galston) 和葛雷 (John Gray)。這不同於羅爾斯和柯亨 (Joshua Cohen) 的「合理多元主義」(reasonable pluralism)，以人類判斷本質的主張來解釋價值觀的差異事實。(Lynch, 2009: 70-71)

宇宙的價值結構提出真理性的說明，主張存在多種分屬不同領域的價值，它們無法化約成共通的尺度來加以比較，因而是不可共量的（incommensurable）。¹¹米提剛（Ross A. Mittiga）指出，當不同價值領域的價值產生衝突時，多元主義論者不訴諸絕對原則來解決，而是根據個別案例的情況來權衡衝突的考量。（2017：412）換言之，對多元主義論者來說，人類不能，也不應該對價值的重要性做出一種適用所有情境的排序。若此，則筆者高度質疑，他們如何會視政治領域的價值「永遠」優先於其他非政治價值。此外，就算同意政治價值具優先性，最多只能支持多元主義論者會同意平等自由權原則，但要推論他們也會認可第二正義原則（特別是差異原則），還必須提供相關理據。這個原則適用於涉及社會和經濟不平等的制度，對這類制度的判斷經常牽涉到如何排序不同價值領域的諸多價值，多元主義論者不會接受所有情境裡的所有相關制度都適用第二正義原則蘊含的價值排序。

筆者認為，更讓人質疑的是典型例子中對「全面性宗教學說」和「古典效益主義」之論述。以下分別探討這兩種學說的支持者會如何看待兩個正義原則，以及他們能否對之形成共識。

（一）「合理的」宗教學說之歧義判準

在重疊共識的論證裡，羅爾斯並未宣稱「所有」全面性宗教學說都會接受「作為公平的正義」。作為典型例子的是「合理的」宗教學說。然而，對於什麼是合理的全面性宗教學說，羅爾斯採取有歧義的判斷標準。這是因為，根據羅爾斯的論述，合理的宗教學說肯定信仰自由，所以它

¹¹ 蓋爾斯敦指出，價值多元主義的一個基本前提是反對價值之間存在一種全面性的排序，即主張沒有任何單一價值或價值的組合是在所有情況下都優先的。換言之，價值多元主義者拒絕「無論在什麼情況且無論要犧牲何種價值，有某些價值永遠優先於其他價值」的觀念。（Galston, 1999：770）

們會同意宗教容忍原則，並支持憲政體制的各種基本自由。這裡預設了「只有肯定信仰自由的宗教學說，才是合理的學說」，問題是，這個假設已超出羅爾斯對合理全面性學說的定義。(Barry, 1995: 909) 前文曾提到，羅爾斯表示自己相當寬鬆地認定合理全面性學說：它符合理論理性和實踐理性的運用，且學說內容穩健地發展。但是在論證重疊共識的典型例子時，他卻加入額外的標準——合理的宗教學說肯定信仰自由，這構成論述上的不一致。¹²並且，一旦加入這個蘊含自由主義價值的標準，而重疊共識僅考慮符合該標準的宗教學說，那麼便排除了所有不接受自由主義價值的學說。

對於上述批評，羅爾斯可能如此回應：他提到的合理全面性學說的三個特徵僅是必要條件，而非充分條件。換言之，並非所有符合三個特徵的學說都是合理的。若此，則可加入額外的認定標準，主張唯有肯定信仰自由的宗教學說才是合理的。¹³問題是，這只能證明合理的宗教學說會支持平等自由權原則，卻無法證明它們必然會支持第二正義原則。許多肯定信仰自由的宗教學說並不認為所有職位都要在機會的公平平等條件下對所有人開放，可能也不同意社會和經濟制度的相關措施必須對社會最不利者最有利。那麼，我們便難以設想這些學說的信仰者如何可能「在其學說內部」找到支持第二正義原則的道德理由。

¹² 陳宜中指出，羅爾斯是在兩個不同層次使用「合理的」一詞，因此製造出許多混淆。在政治層次上，他主張不接受自由主義式政治正義觀的學說是「政治上不合理的」；但在社會層次上，只要滿足多元社會公認的一些信念形成判準，就被視為合理的。而羅爾斯對合理性的論述基本上是為了捍衛自由主義式政治正義觀的正當性。(2001: 62-66)

¹³ 在〈再論公共理性概念〉一文裡，可以找到支持這種回應的論述。羅爾斯指出，所有合理的全面性學說（包括宗教或非宗教的學說）都肯定公民擁有平等的權利和自由，包括良心自由和宗教自由。不支持平等基本自由的學說就不是合理的。(1999a: 608-609)

假設羅爾斯在認定合理學說方面持續保持寬鬆立場，¹⁴不添加額外標準，把合理全面性學說的三個特徵視為充分條件，那麼合理的宗教學說甚至不必然會支持平等自由權原則。貝瑞（Brian Barry）指出，許多符合三個特徵的宗教學說不排除壓制異教徒的信仰，理由是這類信仰可能誘使忠實信徒偏離真實的道路。（1995：909-910）修謨爾（Michael Huemer）也指出，堅信基督教基本教義派（Christian fundamentalism）的人可能主張政府應該為上帝的意志服務，施行基督教的道德價值，虔誠的基督徒也不會支持容忍無神論和違反教義的性行為。（1996：378、388）若此，則無法期待所有符合三個特徵的合理宗教學說都會支持平等自由權原則。

根據以上分析，筆者認為，在全面性宗教學說這個典型例子裡，羅爾斯面臨的難題是：無論是否加入「只有肯定信仰自由的宗教學說，才是合理的學說」這個額外判準，他都無法成功地證明兩個正義原則能夠成為所有全面性宗教學說的共識。

（二）對古典效益主義的不一致論述

更讓人質疑的是羅爾斯有關古典效益主義者可以對「作為公平的正義」形成共識的觀點。因為效益主義的社會正義原則跟「作為公平的正義」的兩個正義原則極不相同，對許多社會制度的正義性將得出對立的結論。若此，則效益主義者如何可能像羅爾斯所說，會基於原則應用的簡單性考量，同意「一種在內容上是自由主義式的政治正義觀是令人滿意的，在把所有事情納入計算後，或許甚至是最好、最可行地近似效益原則的要求」（PL：170）？對此，一種可能的回應是，羅爾斯這裡說的

¹⁴ 埃德蒙森（William A. Edmundson）認為，羅爾斯有好理由對全面性學說的合理性採取寬鬆的認定。把合理性概念拿來當成論證穩定性的工具而加以緊縮，並非他打算要做的事。（2016：546-547）

「自由主義式的政治正義觀」，其內容僅包含「作為公平的正義」的第一正義原則。但就算是這樣，筆者認為，根據羅爾斯前期對古典效益主義的理解，效益主義者不會支持包含平等自由權原則的政治正義觀。

在《正義論》裡，羅爾斯對古典效益主義提出嚴厲的批判。他認為效益主義的正義原則把用來指導個人理性選擇的原則，錯誤地擴大應用在社會整體，因而未能嚴肅看待人與人之間的差異。(TJ：26-27)這是說，當個人面臨相互衝突的欲望時，理性的解決辦法是按照可讓自身整體滿足極大化的方案來行動。個人願意犧牲某些強度較低或當下較小的欲望滿足，以實現強度較高或長期的較大滿足。由於犧牲與獲得滿足的主體是同一人，所以這麼做是理性的。不過，效益主義者在思考社會正義時，卻把所有個人合併為一個整體，進而推論，在個人的決策裡，犧牲某些欲望來成就整體滿足極大化是理性的，那麼社會用多數人的較大滿足來證成犧牲少數個人的欲望滿足也是理性的。(TJ：23-24)由於效益主義忽略被犧牲和獲得滿足的是不同人，所以只要社會整體滿足淨值能增加，它有可能支持侵害個人基本權利和自由的措施。(TJ：450)

在後期理論裡，羅爾斯對古典效益主義的態度有極大的轉變。該學說不再受到嚴厲批判，¹⁵在重疊共識的典型例子裡更被視為盟友——古典效益主義者將會同意以「作為公平的正義」來規範社會主要制度。問題是，假設羅爾斯在《正義論》裡對古典效益主義的描述正確，那麼要如何一致地說明，在理論上可能支持侵犯基本權利的效益主義者，將會放棄效益原則，轉而支持平等自由權原則？這裡同樣有論述不一致的問

¹⁵ 在《政治自由主義》裡，羅爾斯僅在證成原初立場的立約者將會選擇「作為公平的正義」的兩個正義原則時提及，相較於古典效益主義的第一原則（效益原則），兩個正義原則可以對民主社會的自由和平等提供更好的理解。(PL：292)

題。此外，羅爾斯更難說明的是，古典效益主義者如何會支持第二正義原則。因為當社會制度落實機會公平平等原則和差異原則時，很可能造成社會整體欲望滿足淨值減少的結果。由於羅爾斯後期並未表明他要修正《正義論》裡對古典效益主義的理解或放棄對該學說的批判，因此，僅訴諸「原則在應用上的簡單性」，主張古典效益主義者可以對「作為公平的正義」形成共識，缺乏理論上的一致性和說服力。

綜合以上分析，筆者認為，即使「作為公平的正義」被明確定位成政治的正義觀，但它的兩個正義原則無法獲得所有合理全面性學說（特別是全面性宗教學說和古典效益主義）的認可。¹⁶由於典型例子的論證無法成立，因此，重疊共識的論證是失敗的。

二、限縮共識的對象，行得通嗎？

前述的分析顯示，儘管在特定詮釋下，合理全面性學說的支持者可能會同意平等自由權原則，但對於是否應該以第二正義原則（特別是差異原則）規範涉及社會和經濟不平等的制度措施，人們將有深刻的歧見。在1999年《正義論》修訂版的〈導論〉裡，羅爾斯也承認：

¹⁶ 本文的一位審查人指出，在論證重疊共識的可能性時，有另一種理由會影響人們對正義原則的接受程度：人的其中一種身分是公民。做為公民，人們無須訴諸全面性學說，可以直接運用實踐理性，從公共政治文化中理解、應用並接受支持兩個正義原則的理據，由此證成正義原則。筆者認為，這種論證取向採取的是羅爾斯區分三種觀點——原初立場裡立約者的觀點、良序社會裡公民的觀點、我們自己的觀點（point of view of ourselves）——中的第二種觀點，嘗試以之證成重疊共識。然而，這忽略了在羅爾斯的理論裡，採前兩種觀點而排除全面性學說的考量，屬於證成正義原則的第一階段。但重疊共識（證成的第二階段）涉及的是第三種觀點，即「闡述作為公平的正義，且將之當成一種政治正義觀來進行檢驗的你和我」的觀點。（PL：28）羅爾斯指出，這裡要運用反思均衡的檢驗，由個人去確認這種政治正義觀在何種程度上闡述出自己較為堅定的、有關正義的深思熟慮信念。由於羅爾斯提出重疊共識論證的目的，是要證明「作為公平的正義」能贏得各種合理全面性學說支持者的認可，所以從人做為公民的身分，在不考慮自己的全面性學說的情況下接受兩個正義原則，並不符合其欲達成的理論目的，甚至可能根本消除了重疊共識論證在理論上的必要性。

支持平等基本自由的論證乍看之下較為強烈，支持差異原則的論證則包含各種考量的一種較為微妙的平衡……我持續認為差異原則是重要的，還將提出支持它的論點……。但最好能承認這個論點不太明顯，且不太可能具有如同支持兩個優先原則的論證之效力。(1999b: xiv)

如果要挽救重疊共識的論證，那麼或許可以嘗試限縮達成共識的對象，主張各種合理全面性學說能對第一正義原則形成重疊共識。這種修正符合羅爾斯後來以自由主義式的政治正義觀家族 (family of liberal political conceptions of justice) 來說明公共理性 (public reason) 多元形式的觀點。羅爾斯指出，各種自由主義式的政治正義觀都賦予某些基本權利和自由特殊的優先性，且透過一些措施來確保所有公民都能擁有有效運用自由的適當全方位手段。「作為公平的正義」只是這個家族裡的其中一種正義觀。(1999a: 581-582) 我們可合理推論這個家族裡的各種政治正義觀都會接受類似羅爾斯第一正義原則的要求。

許多學者支持這樣的修正。貝瑞指出，要讓自由主義式的民主政治保持穩定，只需要人們同意某種類似平等自由權原則的規範。經驗顯示，當代民主社會能夠容許對第二正義原則涉及的社會和經濟不平等問題，存在衝突的看法。(Barry, 1995: 910) 希爾也認為，如果把「作為公平的正義」視為一種自由主義形式所需的最小觀點，包含有關政府結構和政治程序的原則以及平等的基本權利和自由，但不包括機會公平平等原則和差異原則，那麼這樣的正義觀比較可能成為重疊共識。(Hill, 1994: 347-348)

筆者認為，把「作為公平的正義」的實質內容限縮在平等自由權原則，確實能提高達成重疊共識的可能。然而，羅爾斯仍將面臨兩大困難。

首先，由於共識的對象是各種自由主義式政治正義觀共享的原則，這將使得羅爾斯的政治正義觀沒有可區別於其他正義觀的特點。若此，則要付出的理論代價是，他必須放棄「作為公平的正義」是最穩定的正義觀，所以可以之解決穩定性問題的論證取向，這將破壞他的穩定性論證跟整體理論的關連。其次，能達成共識的第一正義原則僅包含相當有限的政治價值。當人們面對不同的基本權利相互衝突，對特定自由的界限有嚴重歧見，或涉及社會和經濟不平等問題的制度爭議（例如都市更新、基本工時等政策）時，僅依據平等自由權原則將難以做出仲裁。因此，就算限縮後的政治正義觀能成為重疊共識，仍無法維繫多元社會的穩定。

綜合以上所述，可以得出以下結論：重疊共識論證面臨的困難是，如果「作為公平的正義」的內容足夠厚實，包含兩個正義原則時，那麼它很難成為所有合理全面性學說的共識。但如果從它的內容刪除第二正義原則，使之有可能成為共識，那麼羅爾斯必須放棄宣稱它是穩定的正義觀，且過薄的內容無法解決爭議。因此，重疊共識論證不能成功解決多元社會的穩定性問題。

三、正義觀的合理多元事實

我們可以從另一個角度探討重疊共識論證失敗的原因：儘管羅爾斯後期正視到全面性學說的合理多元事實，但卻忽略了人們對於哪種正義觀最適合用來規範社會制度，也有合理的歧見。重疊共識的概念意味著，儘管對於涉及宗教、道德、形上學的終極信念與價值有著分歧主張，人們仍能對相同的政治正義觀達成共識。但事實上，當代社會對「正義觀」和「價值觀」都存在合理多元的事實。巴爾（Kurt Baier）指出，除了「作為公平的正義」，還有至善論（perfectionism）、效益主義、極端自由主義（libertarianism）、社會主義的正義觀，人們不可能對特定正義原則達成

共識。(1989: 775) 誠如沃爾準 (Jeremy Waldron) 所說：「『全面性宗教、哲學和道德學說的多元』不是我們在當代民主社會裡唯一要處理的多元。我們也必須處理正義的多元和正義的歧見。」(1994: 380) 換言之，公民不但抱持不同的全面性學說，也支持不同的政治正義觀，我們無法證明只有「作為公平的正義」是合理的。¹⁷如果更充分地正視當代民主社會的合理多元事實，那麼必須承認合理的政治正義觀也是多元而非單一的。無論「作為公平的正義」還是其他正義觀，都不可能成為重疊共識。

筆者認為，儘管重疊共識的論證避免了前期預設康德的全面性學說，因而跟整體理論肯定合理多元事實不一致的問題，但羅爾斯採取的仍是相同的論證策略：即證明「作為公平的正義」這種正義觀是穩定的，所以依據該正義觀的兩個原則來規範主要制度，便可達成出於正當理由的社會穩定。這可以解釋他為什麼始終如此重視「正義原則的穩定性」。差別僅在於前期是依據「作為公平的正義」跟個人的「好」相互契合，後期則訴諸這種正義觀跟合理全面性學說的契合性（可對之形成共識），來解決穩定性問題。根據以上分析，重疊共識的論證是失敗的，它在理論上可能獨斷地排除非自由主義式的全面性學說，且其正義原則無法成為合理全面性學說的共識。我們可以說，在當代社

¹⁷ 在《正義論》裡，羅爾斯談到，人們共有的正義概念 (concept of justice) 具體表現在決定制度設計問題時，將呈現出不同的正義觀。(TJ: 5-6) 他後期也承認，「除了衝突的全面性學說，政治自由主義確實認知到，在任何真實的政治社會，社會的政治爭論裡有許多不同的自由主義式政治正義觀彼此相互競爭。」(1996: xlvi) 就此而言，他肯定正義觀的多元。然而，若套用羅爾斯的用語，他肯定的只是正義觀的「簡單多元」，而非「合理多元」。因為他前後期理論都主張「作為公平的正義」是穩定的正義觀，且持續為差異原則辯護，批評極端自由主義的分配正義觀。桑德爾 (Michael J. Sandel) 指出，羅爾斯可能宣稱，對分配正義雖有多元的事實，但沒有「合理的」多元事實。桑德爾質疑，羅爾斯的理論對道德和宗教的差異性表現得相當寬容，但奇怪的是，這種容忍態度卻不適用在人們對正義的不同意見上。(1998: 205-206)

會存在「正義觀的合理多元事實」下，任何嘗試從正義原則層次建構的穩定性論證都不可能成功。我們必須另覓他途，為多元社會的穩定性問題提出其他解決辦法。

陸、道德心理的穩定性論證初探

就穩定性問題關切的是如何促使人們出於道德上正當的理由，自願遵守社會規範而言，關鍵在於適當的道德動機。於是，一種可能的論證取向是從「道德心理」的層次來建構穩定性論證。道德心理探討人的道德情感與能力，羅爾斯前後期理論都提出相關主張。¹⁸筆者認為，我們可以從他的道德心理觀點來重構適用於多元社會的穩定性論證。

對羅爾斯來說，維繫社會穩定的最重要道德能力是正義感，它涉及遵守社會規範的有效欲望，也使得個人願意為建立正義的制度而努力。前文指出，羅爾斯前期提出正義感的三個發展階段，把「原則的道德」視為最高階段，這受到心理學家柯爾柏格（Lawrence Kohlberg）理論的影響。柯爾柏格透過道德兩難的問答研究，把道德發展分為六個階段，「普遍倫理原則」是最成熟的階段。¹⁹然而，柯爾柏格的理論有兩方面的問題：首先，從實證面向來看，批評者認為他的階段區分未能真實反應人類的道德心理發展。吉利根（Carol Gilligan）指出，參與柯爾柏格研究

¹⁸ 從道德心理來建構穩定性論證的觀點，獲益於許漢教授，在此表達由衷的感謝。

¹⁹ 柯爾柏格的研究是給受試者看道德困境案例——漢斯的兩難（Heinz's dilemma），根據受試者給出的不同類型理由，把道德發展分成三個層次。每個層次包含兩個階段，越後面的層次和階段表示道德發展越成熟。首先是前習俗層次（preconventional level），包含階段一「他律的道德」和階段二「個人主義、工具性目的和交換」。其次是習俗層次（conventional level），包含階段三「相互的人際期待、關係和人際遵從」和階段四「社會體系和良心」。最後是後習俗層次（postconventional level），包含階段五「社會契約或效益和個人權利」和階段六「普遍倫理原則」。（Kohlberg, 2010: 72-74）

的 84 位受試者都是男性，實驗方法有性別上的偏頗。由此得出獨尊普遍原則的觀點，更貶低了關懷 (care) 和對他人需求的敏感性在道德上的重要價值。其他研究也發現，很少有人能達到柯爾柏格的較高階段。(1982: 18) 若此，則缺乏有力的經驗證據證明，人們可能發展出奠基於原則的有效正義感。其次，從規範理論的面向來看，無論柯爾柏格的「普遍倫理原則」或羅爾斯的「原則的道德」，都屬於正義原則的層次。前文已指出，根據這類觀點來建構穩定性論證，無法成功解決穩定性問題。

筆者認為，由於當代社會存在正義觀的合理多元事實，所以適用這類社會的穩定性論證必須排除對特定正義原則或正義觀的依賴。於是，我們必須修改羅爾斯對「良序社會」的定義——在這樣的社會裡，「每個人都接受相同的正義原則」，且「主要制度滿足這些原則」。(TJ: 4-5; PL: 35) 重構穩定性論證的第一步，是放棄這兩個預設。此外，這個論證涉及的正義感不再是「一種運用並按照正義原則來行動……的有效欲望」。(TJ: 567. 粗體為筆者所加) 它只需要是一種願意從事公平合作，因而傾向遵守公平合作條件的能力。²⁰我們將之稱為「公共合理性」(public reasonableness) 的能力。²¹

以下將嘗試建構一種奠基於道德心理觀點的穩定性論證取向。首先說明公共合理性能力的內涵及其如何形塑維繫社會穩定所需要的道德動機。接著指出，當公民普遍具備這種能力時，可進一步證成「正義」和「好」之間的弱意義契合性，強化個人維持正義感的意願。

²⁰ 布倫南 (Samantha Brennan) 和諾加 (Robert Noggle) 指出，在孩童的道德心理方面，羅爾斯需要提出的是一種包含「合作道德」的發展模式，而這種道德就是正義感的基礎。心理學家柯爾柏格和皮亞傑 (Jean Piaget) 的研究對穩定性問題最重要的部分，是跟公平合作有關的發展論點。(Brennan and Noggle, 2000: 60-61)

²¹ 「公共合理性」一詞出自莫西度 (Stephen Macedo)。(1990: 40)

一、公共合理性的道德動力

我們可以從羅爾斯的道德心理觀點，來說明願意從事公平合作、遵守合作條件的公共合理性能力。他後期同樣對其道德心理觀點做了「政治的」轉向，強調這種架構是用來表達政治性人觀（做為自由平等人的公民）和公民身分（citizenship）的理想，而非源於自然科學探討整體人性的心理學。（PL：86-87）此外也不再談論家庭和結社對個人道德發展的影響。²²羅爾斯主張，理想的公民是「合理的」（reasonable），這包含兩個面向：（1）願意提出可期待他人也會認可的公平合作條件，並且，只要其他人願意遵守這些規範，自己也會遵守。（2）能夠認知到「判斷的負擔」（burdens of judgment）限制了自己能對他人提出的證成，並且在指導憲政體制中政治權力的正當行使時，接受這種負擔造成的結果。（PL：49、54）以下從這兩個面向來說明公共合理性的能力，以及它如何為個人帶來遵守社會規範的道德動機。

首先，公共合理性意味著公民願意提出他人也會認可的社會合作公平條件，這些合作條件展現在用來分配權利、義務和其他社會價值物的各種社會制度裡。「只要其他人願意遵守，自己也會遵守」，則指出遵守合作條件的義務具有相互性。在修正的穩定性論證裡，合作條件不是依據「作為公平的正義」或其他正義觀的原則，而是符合憲政民主政體的

²² 這樣的政治性定位可回應本文一位審查人提出的可能疑慮，即道德心理涉及各種人格特質和心理傾向，於是這種穩定性論證會讓羅爾斯的理論趨近他欲避免訴諸的全面性學說。本文在此探討的是跟穩定性有關的「政治的」道德心理，不涉及特定的全面性學說。此外，筆者認為羅爾斯後期之所以不談家庭（權威的道德）和結社（結社的道德）跟正義感發展的關連，是因為二者的許多理想屬於全面性學說的領域，於是將跟他強調「作為公平的正義」是獨立自恃的政治正義觀產生衝突。希爾也指出，羅爾斯後來放棄前期的心理學法則，是因為它們會涉及一些屬於全面性道德學說的背景條件。（Hill, 2014：211-212）

立法和正當政治程序，且公平對待所有人的規範。²³當然，對於什麼樣的制度才公平，抱持不同正義觀的人們將有分歧的看法。但是就公平合作條件是能得到他人認可的規範而言，合理的公民在思考社會制度時，不能只從自己及所屬團體利益的角度做判斷，必須把其他人和社會整體利益納入考量。此外也不能只因社會制度不符合個人利益，就拒絕遵守相關規範。公共合理性的這部分能力將可減少出於私利的政治歧見，且在一定程度上提高個人遵守社會規範的意願。

其次，公共合理性意味著願意承認判斷的負擔及其結果。在社會制度的爭議裡，就算排除出於私利、不理性或邏輯推理錯誤的因素，公民仍可能持分歧的立場和觀點。羅爾斯以「判斷的負擔」來說明這種合理歧見的根源。它指的是人在做判斷時會受到許多因素的影響，例如，要理性權衡各種目的並評估它們對個人的重要性，要合理評估不同人提出的要求強度以及不同信念和思想體系之運用，都相當困難。此外，用來做判斷的證據經常是複雜或衝突的；人們對相同證據有不同的衡量權重；對許多模糊的概念有不同詮釋；評估證據和衡量價值的方式會受到個人整體生活經驗的影響；爭論雙方有不同種類且帶有不同力量的各種規範性考量，很難做出整體評估；在可實現的各種道德和政治價值之間做取捨時，優先順序的設定和決定都很困難。（PL：56-57）由於沒有人能完全擺脫這些因素的影響，所以即使真心誠意地交換意見、自由討論，公民們仍會對制度提案做出不同判斷，因而導致歧見。就此而言，當公民能夠認知到判斷的負擔及其可能產生合理的歧見，那麼他們將願意承認，自己不同意的立場可能有良好

²³ 這是一種最起碼的底線，類似羅爾斯所謂的憲法共識（constitutional consensus），即社會制度必須符合憲法保障的基本政治權利和自由以及民主程序。但對於這些權利和自由的精確內容和界限，以及哪些進一步的權利和自由被視為基本的，因而應該得到法律的保障，人們仍有歧見。（PL：158-159）

且合理的支持理由，不會只因社會制度不符合自身觀點就拒絕遵守。此外，在討論制度提案時，將願意抱持更開放的心態，嘗試瞭解不同觀點，並且會盡可能蒐集相關資訊和證據，共同找尋更禁得起檢驗的解決辦法。

筆者認為，當正義感被瞭解為包含上述兩個面向的公共合理性能力時，儘管不能完全消除社會制度的爭論，但可減少歧見的部分來源，並開啟公共審議和持續對話的可能。在全面性學說和正義觀的雙重合理多元事實下，要確保公民自願遵守規範，維繫社會穩定，需要的不是訴諸對特殊價值觀或正義觀的「共識」來仲裁歧見，這是不可能達成的任務。如果能促使公民避免訴諸自利考量，願意瞭解不同的立場和觀點，並就制度爭議去尋求公平且彼此都較能接受的解決辦法，將可提高人們接受並遵守最後決議的意願。²⁴若要進一步增強公民遵守社會規範的道德動機，我們還可以在公共合理性能力的基礎上，從道德心理觀點來論證「正義」和「好」的一種弱的契合性。

二、「正義」和「好」的弱契合

本文第參節談到，在前期的契合性論證裡，羅爾斯提出三個理由來論證正義的要求跟個人的「好」相互契合。儘管基於合理多元的事實，必須放棄奠基於康德道德學說的第三個理由，但我們無須連帶拋棄前兩

²⁴ 本文的一位審查人指出，此處的論點似乎傾向透過審議民主的方法來尋求共識，建議補充說明將政治自由主義轉為「透過公民審議尋求政治共識」的模式。筆者的確認為，在合理多元的事實下，透過公民集體審議來處理制度爭議，是比較可欲的作法。然而，筆者並不主張審議的目標在於「尋求政治共識」。因為當代社會許多制度爭議涉及衝突但卻都是合理的價值觀和正義觀，良好的審議仍可能無法讓人們獲得共識。並且，設定以共識作為審議的目標，可能造成忽略或甚至壓制少數觀點的問題。因此，審議的目標僅需設定為此處提到的，「尋求公平且彼此都較能接受的解決辦法」。限於本文篇幅的考量，無法對這個論點做更多說明。有關審議民主模式如何能確保多元社會的穩定，筆者已在另一篇論文裡詳細探討。(2014：24-38)

個理由。因為它們不涉及特定的全面性學說或正義觀，若能做適當補充，則可在一定程度上證成「正義」和「好」的契合性。

首先，對擁有公共合理性能力的公民來說，為個人利益而違反社會規範，必須付出極大的心理代價。這樣的代價除了羅爾斯談到的，因為欺騙和偽善造成個人自發性和自然性的損失外，還必須付出喪失自尊（self-respect）的重大代價。羅爾斯指出，合理的公民希望自己成為且被認為充分合作的社會成員，這支持了他們做為公民的自尊。（PL：81-82）自尊的來源之一是自身價值感，即覺得自己有價值，也確信設定的價值觀和人生計畫值得實行。自尊的重要價值在於，無論個人追求什麼生活計畫和目標，一旦缺乏自尊，就會覺得沒有任何事值得去做，或即使覺得某些事有價值，也會缺乏努力達成它們的意志。（TJ：440；PL：318）人的自身價值感不僅涉及個人的自我認知，也跟他人的肯定有關。當自己這個人和行為得到他人的欣賞和肯定，將傾向對自身價值擁有更多自信。相反地，當個人意識到自己違反社會規範或缺乏合作的德行，將引發道德上的羞恥，這將傾向破壞自我價值感，也會對可能失去他人對自己的尊重而感覺焦慮。（TJ：440-445）更重要的是，公民能否擁有自尊，會受到社會制度的重大影響。假設社會制度剝奪某些人的基本自由，以換取整體較大利益，等於公開認可這些人處於次等地位。這是對他們的極大羞辱，也將因而破壞其自尊。相反地，如果社會制度符合公平合作條件，肯定所有人的平等公民身分，將可支持個人的自尊。（TJ：544-545）²⁵就自尊對個人的重要價值而言，由於做為充分合作成員的自信取決於個人能否展現公共合理性，以贏得自己和他人對自身的肯定，

²⁵ 希爾認為，羅爾斯有關自尊的論點的最大原創性，是他把社會系統對公民自尊的影響作為評估正義理論優劣的判準。（Hill, 2014：206）

且符合公平合作條件的社會制度也有助於維繫自尊，因此，堅持自己的正義感，持續遵守社會規範，跟個人的「好」相互契合。²⁶

其次，根據相互性的概念，參與社會公平合作符合個人的利益。羅爾斯前期把相互性視為正義感三個發展階段的共通特性，後期則把這個概念跟公平合作的道德心理意涵連結起來，我們可以用這個論點來補充前期契合性論證的第二個理由。羅爾斯指出，相互性意味著所有參與合作，根據各種規則和程序的要求履行其職責的人，都將在適當方式下獲益。(PL: 16) 對公民來說，這樣的理性利益展現在透過社會合作，比個人自己或在結社團體裡能達成更全面的價值，且能擴展並維繫每個人想要追求的價值觀。因為人們各有不同的天賦和能力，不可能靠自己完成所有事情，合作可以讓這些天賦和能力互補。而唯有透過公平合作條件來協調彼此的行動和欲追求的目標，個人才可能充分發揮自己的能力，實現其價值觀和目標。(PL: 320-322) 於是，擁有公共合理性能力的公民瞭解到，參與社會公平合作是個人得到各種重要價值的基本前提，社會制度及相關規範確立了共同生活和合作的正當架構，使得多樣的生活計畫、目標和能力得以相互協調，並獲得充分發揮和實現的可能。因此，公民將肯定遵守公平合作的社會規範對個人來說是好的。

以上是從羅爾斯的道德心理觀點出發，去除他的主張裡對特定正義觀的依賴，嘗試重構穩定性論證。儘管由此證成的「正義」和「好」的契合性，要比羅爾斯前期的康德式詮釋或後期的重疊共識來得弱，但卻比較適用於當代的多元社會。²⁷此外，筆者承認這裡提出的弱契合只是

²⁶ 鮑德溫 (Thomas Baldwin) 在正義觀層次提出類似論證。他主張符合相互性的正義觀在制度上的實現，將可提昇公民的自尊，於是可以證成正義和個人的「好」的契合性。(2008: 265-269) 本文提出的論證不依賴特定正義觀，在論證基礎上有別於鮑德溫的主張。

²⁷ 本文的一位審查人建議把道德心理論證視為重疊共識論證在穩定性問題的補充，而非取代，

理論上的推論。若要讓它實際內化為個人持續遵守規範的有效道德動機，還必須透過社會制度和教育措施，培養公民普遍擁有公共合理性的正義感能力，以及深刻認識「正義」和「好」的契合性。如此才能讓這種奠基於道德心理的穩定性論證成為應用上可行的。²⁸因此，修正後的穩定性論證還需要補充民主公民教育的概念性觀點和實作措施。筆者在此提出的僅是初步的修正方向，要如何系統性地結合道德心理觀點和公民教育，建構完整的穩定性論證，仍有待未來進一步探究。

柒、結論

羅爾斯表示，政治自由主義希望回答的問題是：「由自由平等的公民——他們因各種合理的宗教、哲學與道德學說而保持著深刻的分化——組成的正義且穩定的社會，如何可能長久存續？」(PL:4) 這種對穩定正義社會的關懷，貫穿了羅爾斯前後期的學說，也說明了他持續修正理

因為兩種論證都是羅爾斯論述穩定性的理論，只談任一者都不完整。對此，必須澄清的是，之所以要重構道德心理論證，是因為根據前述分析，重疊共識的論證無法成立。因此，根據道德心理觀點所重構的穩定性論證無意作為重疊共識論證的補充。

對本文提出的道德心理論證的一種質疑是，這種論證取向沒有真正回答羅爾斯探討穩定性時想要處理的問題。本文的另一位審查人認為，羅爾斯希望能證成在他定義的良序社會裡，我是否應該總是正義地行動？」而他的問題意識聚焦於「正義原則的穩定性」，嘗試探討在一個受其正義原則規範的良序社會裡，人們在理性上是否有足夠的理由，賦予正義感優先性，而正義感的內容必然預設一組特定的正義原則。對此，筆者同意審查人所說，穩定性問題的關鍵在於正義感的優先性。羅爾斯希望能證成在他定義的良序社會裡，兩個正義原則的穩定性將促使人們賦予正義感優先於其他動機的地位，因而願意總是依照正義原則來行動。但問題是，羅爾斯表明良序社會是存在合理多元事實的憲政民主社會，而本文的分析指出，這類社會裡的人們不會認可相同的正義原則。筆者認為，因為羅爾斯的理論整體目標是要回答「多元」社會的穩定性問題，所以必須放棄從正義原則的穩定性來解決社會穩定問題之論證取向。就此而言，這裡提出的道德心理論證確實不是以羅爾斯的問題意識來回答穩定性問題，而是在獨立於特定正義觀或正義原則的基礎上，去論證正義感優先的理據。

²⁸ 埃德蒙森指出，羅爾斯對良序社會的教育只給出模糊的描述。然而在考量跨世代的社會穩定時，唯有透過道德和公民教育，讓年輕人培養出正義感，才可能達成出於正當理由的穩定。(Edmundson, 2016: 550-552)

論的原因。²⁹儘管後期正視到當代社會存在多樣的合理全面性學說，因而重新定位整體理論並做出許多修正，但誠如蓋爾薩夫（Jon Garthoff）指出的，羅爾斯採取的穩定觀、根據正義觀的穩定性來評估其優劣，以及在穩定性問題上偏愛「作為公平的正義」這幾方面仍沒有改變。（2016：297）在前期，羅爾斯主張兩個正義原則能促使人們發展出強烈的正義感，且正義原則的要求符合個人的「好」，所以是穩定的正義觀。到了後期，他主張在全面性學說的合理多元事實下，「作為公平的正義」能成為所有合理學說的重疊共識，所以是最穩定的正義觀。人們將基於對這種政治正義觀的共識，自願依據兩個正義原則來行動，遵守符合正義原則的社會制度。

本文指出，羅爾斯之所以放棄前期的契合性論證，以重疊共識的論證取代，是為了避免訴諸康德的全面性學說來證成「作為公平的正義」的穩定性，跟整體理論肯定合理多元事實的觀點不一致的問題。儘管重疊共識的論證一致於羅爾斯後期理論的政治性定位，但「作為公平的正義」無法成為重疊共識。我們不應該忽略合理的政治正義觀也是多元的，因此，任何依賴特定正義觀的穩定性論證都無法成立。

面對多元社會的穩定性問題，筆者提出一種奠基於道德心理觀點的論證取向。儘管這樣的論證只能提供「正義」和「好」的較弱契合性，無法充分保證公民能始終堅持遵守社會規範。但這是政治哲學充分正視社會的多元性所必須承受的理論負擔。因為在正義觀的合理多元事實下，不可能有任何正義觀能得到所有公民的認可。本文指出，若能肯定作為公共合理性的正義感能力在維繫社會穩定方面發揮的作用，並致力

²⁹ 許多學者都指出，對穩定性問題的關切以及對契合性論證的自我批判，促使羅爾斯做出理論轉向。參見 Barry（1995：880）；Lynch（2009：72）；Weithman（2011：5）；Freeman（2003：278）；Kukathas and Pettit（1990：141）；周保松（2013：204）。

透過教育和相關社會制度來培養公民普遍具備這種道德能力，仍可在一定程度上維繫正義社會的長期穩定。

參考文獻

- 吳澤玖 (2014)。〈審議民主與多元社會的穩定〉，《政治與社會哲學評論》，49：1-58。DOI: 10.6523/168451532014060049001
- 周保松 (2013)。《自由人的平等政治 (增訂版)》，北京：生活讀書新知三聯書店。
- 陳宜中 (2001)。〈羅爾斯與政治哲學的實際任務〉，《政治科學論叢》，14：47-74。
- Baier, Kurt (1989). "Justice and the Aims of Political Philosophy." *Ethics*, 99, 4: 771-790. DOI: 10.1086/293121
- Baldwin, Thomas (2008). "Rawls and Moral Psychology." Russ Shafer-Landau (ed.). *Oxford Studies in Metaethics*, vol. 3 (247-270). New York: Oxford University Press. 2008.
- Barry, Brian (1995). "John Rawls and the Search for Stability." *Ethics*, 105, 4: 874-915. DOI: 10.1086/293756
- Brennan, Samantha and Noggle, Robert (2000). "Rawls's Neglected Childhood: Reflections on the Original Position, Stability, and the Child's Sense of Justice." Victoria Davion and Clark Wolf (eds.). *The Idea of a Political Liberalism: Essay on Rawls* (46-72). Maryland: Rowman & Littlefield Publishers, Inc.
- Cohen, Gerald Allan (2008). *Rescuing Justice and Equality*. Cambridge, Mass.: Harvard University Press. DOI: 10.4159/9780674029651
- Edmundson, William A. (2016). "Coercion, Stability, and Indoctrination in the Pejorative Sense." *Jurisprudence*, 7, 3: 540-556. DOI: 10.1080/20403313.2016.1236893

- Freeman, Samuel (2003). "Congruence and the Good of Justice." Samuel Freeman (ed.). *The Cambridge Companion to Rawls* (277-315). Cambridge, Mass.: Cambridge University Press.
- Galston, William A. (1999). "Value Pluralism and Liberal Political Theory." *American Political Science Review*, 93, 4: 769-778. DOI: 10.2307/2586111
- Garthoff, Jon (2016). "Rawlsian Stability." *Res Publica*, 22, 3: 285-299. DOI: 10.1007/s11158-015-9292-z
- Gilligan, Carol (1982). *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, Mass.: Harvard University Press.
- Habermas, Jürgen (1995). "Reconciliation through the Public Use of Reason: Remarks on John Rawls's Political Liberalism." *The Journal of Philosophy*, 92, 3: 109-131. DOI: 10.2307/2940842
- Hill, Thomas E. Jr (1994). "The Stability Problem in Political Liberalism." *Pacific Philosophical Quarterly*, 75: 333-352. DOI: 10.1111/j.1468-0114.1994.tb00134.x
- (2014). "Stability, a Sense of Justice, and Self-Respect." Jon Mandle and David A. Reidy (eds.). *Companion to Rawls* (200-215). Chichester, West Sussex: Wiley Blackwell. DOI: 10.1002/9781118328460.ch11
- Hobbes, Thomas (1998). J. C. A. Gaskin (ed. and intro.) *Leviathan*. Oxford; New York: Oxford University Press.
- Huemer, Michael (1996). "Rawls's Problem of Stability." *Social Theory and Practice*, 22, 3: 375-395. DOI: 10.5840/soctheorpract199622311
- Kohlberg, Lawrence (2010). "Resolving Moral Conflicts within the Just Community." Carol Gibb Harding (ed.). *Moral Dilemmas and Ethical Reasoning* (71-97). New Brunswick: Transaction Publishers.
- Krasnoff, Larry (1998). "Consensus, Stability, and Normativity in Rawls's Political

- Liberalism.” *The Journal of Philosophy*, 95, 6: 269-292. DOI: 10.5840/jphil199895613
- (2015) “Stability.” Jon Mandle, SUNY Albany and David A. Reidy, University of Tennessee, Knoxville (eds.). *The Cambridge Rawls Lexicon* (804-810). Cambridge, United Kingdom: Cambridge University Press. DOI: doi.org/10.1017/CBO9781139026741.212
- Kukathas, Chandran and Pettit, Philip (1990). *Rawls: A Theory of Justice and Its Critics*. Cambridge: Polity Press.
- Lister, Andrew (2015). “Congruence.” Jon Mandle, SUNY Albany and David A. Reidy, University of Tennessee, Knoxville (eds.). *The Cambridge Rawls Lexicon* (133-138). Cambridge, United Kingdom: Cambridge University Press. DOI: 10.1017/CBO9781139026741.043
- Lynch, Sterling (2009). “The Fact of Diversity and Reasonable Pluralism.” *Journal of Moral Philosophy*, 6, 1: 70-93. DOI: 10.1163/174552409X365937
- Macedo, Stephen (1990). *Liberal Virtues: Citizenship, Virtue and Community in Liberal Constitutionalism*. Oxford: Clarendon Paperbacks.
- Mittiga, Ross A. (2017). “Pluralist Partially Comprehensive Doctrines, Moral Motivation, and the Problem of Stability.” *Res Publica*, 23, 4: 409-429. DOI: 10.1007/s11158-016-9335-0
- Rawls, John (1971). *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.
- (1993). *Political Liberalism*. New York: Columbia University Press.
- (1995). “Reply to Habermas.” *The Journal of Philosophy*, 92, 3: 132-180. DOI: 10.5840/jphil199592336
- (1996). “Introduction to the Paperback Edition.” *Political Liberalism* (xxxvii

- lxii). New York: Columbia University Press.
- (1999a). "The Idea of Public Reason Revisited." Samuel Freeman (ed.). *John Rawls: Collected Papers* (573-615). Cambridge, Mass.: Harvard University Press.
- (1999b). *A Theory of Justice* (Revised Edition). Oxford: Oxford University Press.
- Sandel, Michael J. (1998). "A Response to Rawls' Political Liberalism." *Liberalism and the Limits of Justice* (184-218). Mass.: Cambridge University Press. DOI: 10.1017/CBO9780511810152.008
- Waldron, Jeremy (1994). "Disagreements about Justice." *Pacific Philosophical Quarterly*, 75: 372-387. DOI: 10.1111/j.1468-0114.1994.tb00136.x
- Weddle, Aaron (2017). "On the Ideal of Political Stability." *Dialogue*, 59, 2-3: 217-225.
- Weithman, Paul J. (2011). *Why Political Liberalism?: On John Rawls's Political Turn*. New York; Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780195393033.001.0001

