The Arrow Impossibility Theorem

选择的悖论

阿罗不可能定理与社会选择真相

[美] 埃里克·马斯金 [印] 阿马蒂亚·森 等◎著 黄永◎译

中信出版集团 · CHINACITICPRESS

The **Arrow**
Impossibility
Theorem

选择的悖论

阿罗不可能定理与社会选择真相

埃里克·马斯金（ERIC MASKIN）
阿马蒂亚·森（AMARTYA SEN）
肯尼斯·阿罗（Kenneth J. Arrow）
帕萨·达斯古普塔（Partha Dasgupta）◎著
普拉桑塔·帕坦尼克（Prasanta K. Pattanaik）
约瑟夫·斯蒂格利茨（Joseph E. Stiglitz）

黄永◎译

中信出版集团 · CHINACITICPRESS · 北京

图书在版编目（CIP）数据

选择的悖论 /（美）马斯金，（印）森　等著；黄永译
. -- 北京：中信出版社，2016.6
（大师公开课系列）
书名原文：The Arrow Impossibility Theorem
ISBN 978-7-5086-6033-2

I. ①选… II. ①马… ②森… ③黄… III. ①经济学
－通俗读物 IV. ① F0－49

中国版本图书馆 CIP 数据核字（2016）第 057940 号

选择的悖论：阿罗不可能定理与社会选择真相

著　　者：[美] 埃里克·马斯金　[印] 阿马蒂亚·森　等
译　　者：黄　永
策划推广：中信出版社（China CITIC Press）
出版发行：中信出版集团股份有限公司
（北京市朝阳区惠新东街甲 4 号富盛大厦 2 座　邮编　100029）
（CITIC Publishing Group）
承 印 者：北京画中画印刷有限公司

开　　本：880mm×1230mm　1/32　　印　　张：5.75　　字　　数：85 千字
版　　次：2016 年 6 月第 1 版　　印　　次：2016 年 6 月第 1 次印刷
京权图字：01-2014-4203　　广告经营许可证：京朝工商广字第 8087 号
书　　号：ISBN 978-7-5086-6033-2
定　　价：49.00 元

阿罗不可能定理与社会选择

肯尼斯·J. 阿罗不可能定理（1950，1951）发表60多年来，已对所有关心社会选择和福利学领域问题的人士产生了深远的影响，而埃里克·马斯金和阿马蒂亚·森沿袭阿罗定理之路，为该领域做出了大量重要贡献。我在研究生期间聆听过阿马蒂亚·森讲授阿罗不可能定理（以及大部分福利经济学的内容），与埃里克·马斯金也相识已久，因此，能够为这本埃里克·马斯金和阿马蒂亚·森在哥伦比亚大学讲授阿罗不可能定理的讲座集撰写导

言，我感到格外荣幸。

阿罗不可能定理发表时代的福利经济学和社会选择理论

阿罗不可能定理几乎改变了所有相关学科和学科分支，是对思维领域罕有的一大贡献。阿罗不可能定理对政治哲学和政治理论亦有着深远的影响，不过福利经济学和社会选择理论仍是该定理所影响的主要领域。在此，很有必要对阿罗发表其著名定理时福利经济学的发展情况进行简要回顾。柏格森（Bergson，1938）和萨缪尔森（Samuelson，1947）对准确定义社会福利函数做出了重要贡献，他们使人们充分认识到，如果经济学家想参与社会政策决策或者评估社会状态，需要做出价值判断，然而福利经济学家在20世纪30年代至40年代关注的价值判断太少。得益于消费者行为理论中序数分析的兴起，以及对于个体之间效用比较可能性的广泛质疑，之前的功利主义传统被摒弃。大部分福利经济学研究的重点集中于帕累托法则（Pareto Principle），即如果社会中所有人绝对偏好某种社会状态x，而非另一种社会状态y，则对整个社会x绝对优于y。[2] 当然，

帕累托法则存在一个重要问题，若两个个体对两种社会状态的偏好排序存在差异，帕累托法则就无法对这两种社会状态做出比较。为了弥补帕累托法则在比较时产生的诸多漏洞，卡尔多（Kaldor，1939）和希克斯（Hicks，1939）引入了“补偿标准”（compensation criterion），但人们很快发现，补偿标准的吸引力不仅在伦理上有限，有时甚至可能造成相互矛盾的结果，比如，对整个社会而言，可能社会状态 x 优于社会状态 y，同时社会状态 y 也优于社会状态 x。阿罗不可能定理发表之前的许多年间，以投票进行社会决策[3]方面的文献出现了一些进展[4]，尤以鲍恩（Bowen，1943）和布莱克（Black，1948a）的文献最为重要。尽管这些论文都发表于顶级经济学杂志，但总体来说，在阿罗定理激发其对投票程序的兴趣之前，福利经济学家并不怎么关注这些论文。

阿罗不可能定理和对偏好加总问题的解释

令 X 为社会中所有可能状态或形势的集合。在任意给定时刻，集合 A 是实际可得的社会状态，是集合 X 的非空子集（可能是真子集）。假设每一个体对 X 都有其偏好排

序。阿罗不可能定理解决了确定 X 中社会状态的社会排序 R 这一问题。确定社会排序 R 的目的在于将其作为社会选择的基础：给定任意可得社会状态的集合 A，社会从中选择一个状态，使其在所有属于 A 的社会状态中 R 的排序最高。

阿罗将其理论建立在如下直觉之上，R 一定基于个人对 X 的偏好排序情形（社会中每一个体均恰好有一个偏好排序）。正如森（2014）所指出的，该直觉显然来自社会选择的民主传统，这也是阿罗正式定义社会福利函数的基石。阿罗的社会福利函数的定义域为一系列个人排序情形，值域为一系列社会排序。因此，对定义域中每一个人的排序情形，一个社会福利函数恰好确定一个对 X 的社会排序。对于如何在已知个人排序的条件下确定 X 中状态的社会排序这一伦理问题，社会福利函数的定义提供了正规的分析框架。除非对社会福利函数施加特定约束或性质限定，否则这一框架不太具有吸引力。但要注意，从直觉上看，社会福利函数定义本身就暗含了基数个人偏好和偏好满足程度的个体间比较对于社会排序的决定不存在任何影响。根据定义，对于定义域中的每一个体的偏好排序情形，社会福利函数恰好确定了一个社会排序。因此，从直觉上看，如果个人对社会状态的偏好排序不变，那么不管个人

偏好的基数信息发生什么变化，社会排序都将不变。阿罗提出了社会福利函数的若干性质。这里我采用森（2014）对这些性质的改写版本。[5] 阿罗提出的条件之一是，社会福利函数的定义域应该是所有逻辑上可能的个人排序情形的集合。给定社会福利函数的定义，该条件等价于要求对于每一逻辑上可能的个人排序，社会福利函数应该确定唯一的社会排序——这一要求被森（2014）称为“无限制定义域（unrestricted domain）公理”。阿罗要求社会福利函数满足福利经济学中广受认可的帕累托法则。他还要求社会福利函数具有另外两个性质，分别是“无关备择项的独立性”（independence of irrelevant alternatives）和“非独裁性”（nondictatorship）。阿罗不可能定理表明，如果 X 包含至少三种不同的社会状态，且社会中的个体数目确定，则不存在同时满足定义域无限制、帕累托法则、无关备择项的独立性，以及非独裁性条件的社会福利函数。由于至少存在三种不同的社会状态，社会中个体数目确定的限定条件并非严格约束，而且四项条件表面上都合理，阿罗不可能定理就像似非而是的悖论。阿罗不可能定理证明，一些似是而非的假定或公理将推导出逻辑上的矛盾，而这类似非而是的结论则有其实用性和重要性，它们通常包含在解决矛

盾给智力带来的挑战之中，包含在寻找为何应该摒弃或修改公理的严谨检视之中。就阿罗不可能定理而言，这项挑战十分巨大，但也促使学者们进行了大量丰富的研究，在不同领域继续推进。

与阿罗不可能定理相关的一个重要问题是社会福利函数的直观解释，而这又与社会福利函数定义中所涉及的对社会排序 R 和个人偏好的解释密切相关。

我们至少可以思考对社会排序的两种不同解释。第一种解释，我们可以将 R 看作当社会依照某一程序或规则进行社会状态排序以做出决策时所得到的比较结果；从这个意义上来说，社会福利函数只是在社会状态排序时所采用的一种决策程序。第二种解释，R 也可被看作是某一个人对社会福利判断的反应，也就是某一个人关于社会状态相对优劣的道德判断；这个人可能是社会中的一员，可能是中央计划者，也可能来自社会之外。这两种解释大不相同：这个人可能会同意采用社会的决策规则，即便这时得到的对两种社会状态 x 和 y 的排序与依照其自身社会福利判断得到的排序有所不同。两种对社会排序解释的差异是某些对阿罗不可能定理的早期评价 [可参见利特尔（Little），1952；柏格森，1954] 中的核心内容[6]，并由森（1977，

2011）进行了详细讨论。阿罗（《社会选择与个人价值》英文版第二版，1963，第 106 页）承认两种解释存在差异[7]，并且支持第一种解释，因为他认为社会的最终选择在根本上取决于采用某种决策程序对个人偏好排序进行加总所得的结果。[8, 9]

与社会排序类似，我们对个人偏好也存在不同的解释。森（2013）阐述了三种对个人偏好的不同解读：通过投票表达的偏好，反映个人利益的偏好，以及反映道德判断的偏好。对于阿罗来说，个人偏好反映了“他认为相关的一切标准”（阿罗，1951，第 7 页）。阿罗（1951）对个人偏好的解释在许多方面都与投票紧密相连。投票者的选择基于其认为相关的一切因素：尽管这也许并不反映其排他性的个人利益 / 福利，也不反映其排他性的道德标准。

较之其他学者，森让我们更好地认识到，社会排序和个人偏好不同解释间的差异十分重要（森，1977，2011，2013）。把社会排序的不同解释与个人偏好的不同观点进行组合，将带来不同的偏好加总问题[10]，尽管其在形式上有所类似，直观上看却迥然不同。显然，阿罗不可能定理的有效性不依赖任何对其整体框架的特定解释，但其框架及定理的吸引力则可能显著依赖于特定的解释。此时，如

果把个人偏好理解为其对社会状态的道德评价，并考虑如何对其进行加总得到社会排序以作为社会决策的基础，那么阿罗对社会福利函数的定义就会因其排除了基数个人偏好和偏好满足程度的个人间比较等特征而看似约束力较低。目前尚不清楚当加总个人道德判断以做出社会决策时，比较对象是个人对某一社会状态的道德优越性的认同程度，还是个人的“道德满足”（moral satisfaction）程度。然而，当把个人偏好理解为其各自的利益水平，同时在对偏好进行加总时加入道德评价标准，从而达到根据福利水平对社会状态做出评价的目的时，情况则大不相同；我们认为此时对个人间利益水平的比较至关重要。

森（2014）强调了阿罗对投票理论中民主传统的应用。这或许有助于区分民主的两大方面。一方面是民主协商的阶段，在这一阶段，会对备择社会状态的个人判断进行商讨和辩论；个人间的利益比较通常是这些判断形成的重要因素。作为这一阶段商讨和辩论[也被阿罗于2014年在其评论中称为“交谈与对话”（conversation and dialogue）]的结果，个人最初对社会状态的判断可能变化，也可能不变。但社会总是需要做出最终决策的。那时，将通过投票对判断进行加总，因为它们是第一阶段即对个人判断的商

讨阶段的成果。阿罗对其理论的解释基于民主决策的后一阶段；因此，其正式框架中并不包含基数偏好和偏好满足程度的个人间比较也就不足为奇了。

对阿罗不可能定理的一些回应

可以想见，对阿罗不可能定理存在两种完全不同的回应。第一种，检视阿罗的整体分析框架和假设条件，至少在偏好加总问题的某些特定解释下，确定是否存在特例去修正它们，进而得到偏离不可能定理结果的可能路径。第二种，认同整体分析框架及假设条件，认为定理是合理的，至少在某些社会选择问题的解释上是合理的，但在下述方面提出异议：既然所有偏好加总规则都有缺陷，都无法满足全部合理条件，那就需要进一步对这些存在缺陷的偏好加总规则的特性进行研究，同时观察其中是否存在某些规则可能优于其他规则的情况。

我们简要回顾一下对阿罗不可能定理第一种回应的例子。我们已经考虑了在偏好加总问题的某些特定解释下，承认基数个人偏好和偏好满意程度的个体间比较更可行，而阿罗对社会福利函数的定义将这一点排除在外了。在这

种情况下，需要同时放松阿罗的无关备择项的独立性假定。即使将社会福利函数的定义调整为允许基数个人偏好和偏好满意程度的个体间比较，通过使任意两种社会状态排序仅依赖于每一个体对这些社会状态的两两排序，阿罗的无关备择项独立性依然会将所有个体间比较排除在外。

但是，还有对个人偏好加总问题的其他解释：这种解释认为无法进行个体间偏好满意程度比较并非不合理。考虑这样一种情形：个体偏好是对社会状态的决定或者“投票”，而社会排序仅仅是社会应用决策过程的结果。让我们在这样的情境下考虑阿罗的条件。由于帕累托法则和非独裁性看上去是公平合理的限制条件，对偏离阿罗不可能结果的尝试通常主要讨论其他条件的合理性。设定无关备择项的独立性似乎主要是出于方便的考虑：如果这个条件不满足，那么想要比较任意两个社会状态，还需要个人对其他（“无关”）社会状态的偏好信息，对个人偏好信息提出这样的要求太过严格。

无限制定义域和社会福利函数的定义结合在一起考虑，意味着对每一个逻辑上可行的个人排序情形都会有一个社会排序。阿罗设定社会排序一定存在这一条件的目的在于（即，一个二元的社会弱偏好关系满足反身性、连通性和

传递性），如果社会弱偏好关系是反身的、连通的和可传递的，那么对每一个有限的可行备择项的集合，都将定义出最优选项。[11] 很多投票规则，比如简单过半数投票（the simple majority rule），在加总个人偏好排序后，会因出现严格社会偏好的循环，导致无法从一些有限的可行社会状态集中确定一个最优的社会状态。如果比较社会状态的目的在于从不同的可行社会状态构成的集合中做出选择，在社会弱偏好关系无法从一些选项集合中确定最优社会选项时，这一目的便无法达成。在阿罗（1950，1951）之后的研究中，这一点很快被揭示，当一个社会排序对每一个有限的可行选项集合都确定了一个最优项时，传递性不是社会弱偏好对每一个有限的可行选项集合都得出最优项的必要条件。很多文章探讨了能否通过放松传递性的要求使结果偏离阿罗不可能定理。但是证明发现，即使是比社会弱偏好关系中的传递性更弱的条件，依然与偏好加总规则中其他希望保留的属性是不相容的。在放松社会弱偏好关系为较弱条件的研究中，最著名的两个不可能结果分别来自吉巴德（Gibbard，1969）和森（1970，a，b），前者将社会弱偏好关系中的传递性放松为社会严格偏好关系中的传递性，后者仅假设社会严格偏好关系的非循环性——这一性质比

社会严格偏好关系的传递性更弱。还有其他令人信服的理由能解释为何需要假设，在给定的个体偏好排序下，每一个可能存在的可行社会状态集对应的社会选择必须是基于社会状态集的全集上定义的固定的社会弱偏好关系吗？给定个人排序，社会选择必须基于定义在 X 上、固定的二元社会弱偏好关系确定这一条件，强化了社会选择中特定的“一致性”。森（1993）使用社会选择而非社会弱偏好关系作为基本概念，论述了阿罗不可能定理的一个版本可以在不要求社会选择满足一致性条件的情况下得到证明。同时可以认为社会选择的一致性与一些我们的直观感受是相冲突的。因此，若一个人同意约翰·斯图亚特·穆勒（John Stuart Mill，1859）的观点，即个人应在与其自身“个人生活”相关的事务上被赋予做决定的自由，那么很容易证明由这个个体根据其个人决策过程得到的社会状态将在一些时候违背一致性，即使是在文献中讨论过的最弱的社会选择的一致性。萨格登（Sugden，1985）给出了一个与已婚夫妇做选择相关的有趣例子，例子中“社会选择”是一个男人对一个女人自由选择求婚或不求婚以及一个女人自由选择接受或不接受一个男人的求婚的结果，这里的社会选择就违背了最弱的选择一致性。

一些文献对阿罗无限制定义域条件的另一方面进行了讨论，其中马斯金（2014）的研究引起了我们的注意。[12] 无限制定义域这一条件要求，对每一个可能的个人排序情形，社会弱偏好关系应当是一个排序。但是如果一些个人排序情形不会出现呢？我们知道简单过半数原则满足帕累托原则、无关备择项的独立性和非独裁条件。如果简单过半数规则，仅在一些“不可能”的个人排序的情形下，会出现社会严格偏好循环，进而无法从一些可行备择项集中选出胜出者，那么我们也许不必太过焦虑。正如马斯金（2014）所写：“这种情形下阿罗不可能定理过于悲观了，如果能通过一个合理且可行的方法对排位次序进行限制，那么五条公理将不再是整体不一致的。”我们知道一些对个人排序情形的限制条件，这些限制条件在一些背景下貌似是可信的，而且这些条件能排除过半数规则的投票循环情形。例如，假设社会选择是选出执政党，所有投票者对这些党派有一个共同的排序方式，排序的标准是这些党派是何种程度的右倾或左倾，并且党派是右派还是左派是投票者唯一关注的标准。我们还知道个人偏好排序满足布莱克（1958，第 7—10 页）提出的“单峰性”。单峰性排除了简单过半数规则下社会严格偏好关系形成循环的情形。但是，

一旦我们考虑的问题超出了投票者仅依据单一标准进行选择时，将很难提出看似可行的限制条件对偏好情形加以限制，以排除简单的过半数规则中可能出现的循环问题。实际上，克莱默（Kramer，1973）的研究得到的精彩而深入的结果证明，在一般情形下，当投票者考虑多方面因素做出选择时，文献中讨论过的、对个人偏好排序情形的大多数限制条件，都不太可能实现。

注意到阿罗公理的一个含义是非常重要的，森（1977，2011，2014）在这方面投入了持续的精力，并且这一点不仅与阿罗的某一条公理相关，而且与无限制定义域、帕累托规则以及无关备择项的独立性同时相关。这三条公理合在一起，在对社会状态进行排序的决策过程中，排除掉了考虑个人偏好信息以外其他任何信息（例如，关于所考虑的社会状态的信息）的可能性。森（2014）给出了令人信服的证明，通过其自身，建立一个社会福利函数的限制性特征，而这一点与穆勒（1859）提出的观点矛盾。穆勒认为一个人有权对与他“个人生活”相关的事务做出自己的选择，无论其他人对他的选择抱有怎样的看法。即使我们将社会福利函数的定义改为引入基数个人偏好和偏好满意度的个体间比较，这一结论依然成立。继森（1970a，