

想象的力量： 从海萨尼转换到无知之幕

曹志刚 北京交通大学

人类不同于其他动物的基本特征之一是能够进行想象，可以对看不见、没有发生乃至不会发生的事情进行思考。比如我们视觉系统的运作很大程度上要靠想象：可以只处理部分信息就“脑补”出完整的图像；比如时间是一个想象的概念：时间并非独立于空间的客观存在，而是辅助我们理解和度量运动的一种想象；比如国家是一个想象的共同体，自我也很可能只是一个想象的概念；比如经济学借助亚当斯密“看不见的手”的深刻洞见和绝妙比喻理解市场运行，供求曲线等经济学基本概念其实也都是想象的产物而不是对现实的描摹。

想象的力量之强大，远超多数人的想象。

本文主要借助博弈论里的一个基本理论——海萨尼转化，以及政治学里的一个重要比喻——无知之幕，从“无知”的视角谈谈想象的力量。学而不思则罔，解决无知的途径不仅仅是观察和学习，还需要充分内省。本文将特别论述的是，有时候可以把多个——通常是无穷多个——问题

打包成一个抽象的问题，使得原本知道的信息失效，通过假装无知的办法后退一步解决一些棘手的重大难题。毫无疑问，想象在这种内省思考中起基本作用。

一、纳什均衡：想象中的稳定

关于博弈即互动决策问题的讨论，在前冯·诺依曼时代更多是一种不成熟的心理学和玄学，是实用主义者发明创造诡计的过程。理性人假设、均衡概念的提炼以及数学形式化语言的应用使得博弈论变成了一门科学。

但是我们也不应被形式化数学符号的表象所蒙蔽。理解博弈论的核心概念、发展脉络以及底层方法论，离不开它脱胎于不成熟心理学的历史以及博弈论是决策科学而非数学或自然科学的基本事实。

在就使用何种均衡概念达成共识前，博弈并非一个良好定义的数学问题。博弈问题的解决首先在于如何定义它的“解”，即我们想得到什么输

出。而博弈问题的输入是对博弈形势客观认知的各种信息。信息是决策的前提,是决策问题的输入。完全确定世界里的决策问题某种意义上就是优化,而博弈跟优化或单人决策相比的本质区别不仅在于彼此的决策相互影响,更在于彼此不了解对手如何出招:一旦清楚了对手如何出招,博弈问题就退化成了普通的决策乃至优化问题。所以在现实世界的各种博弈中,千方百计刺探对手想法是获胜的关键。这些今天看来理论上并不成熟的前冯·诺依曼时代的思想在博弈实践中依然发挥重要作用。

博弈论,无论研究的是完全信息博弈还是不完全信息博弈,都是在某种意义上处理信息缺失的互动决策问题。从信息的视角来看,博弈论的第一个重大突破是在相互不知道彼此选择的时候猜测出一个结果。该猜测出来的结果不仅具有某种稳定性,任何一方都没有激励(不意味着没有能力)单方面改变这种局面(即互为最优),还具有自我应验、无后悔性以及不担心泄密等良好性质。我们知道这个突破性概念就是纳什均衡,一个在20世纪可以跟DNA双螺旋结构的发现相媲美的伟大成就。尽管冯·诺依曼的极大极小解也解决了类似的信息问题,但是由于零和博弈的普适性相对要差一些,所以上述突破一般更多归功于纳什。

无论纳什均衡多么伟大,它对人类行为的约束都只是一种软约束:任何一方没有激励单方面改变纳什均衡并不意味着没有能力改变,也不意味着联合决策没有激励改变纳什均衡。当然,这种软约束跟所有的社会发展规律一样也有一定程度的客观性,对其不尊重也依然会受到相应惩罚。

刘慈欣的科幻小说《三体》里有一个精彩桥段:物理学家观察到前沿物理实验结果后纷纷自杀,因为这与他们深信不疑的科学规律完全相违

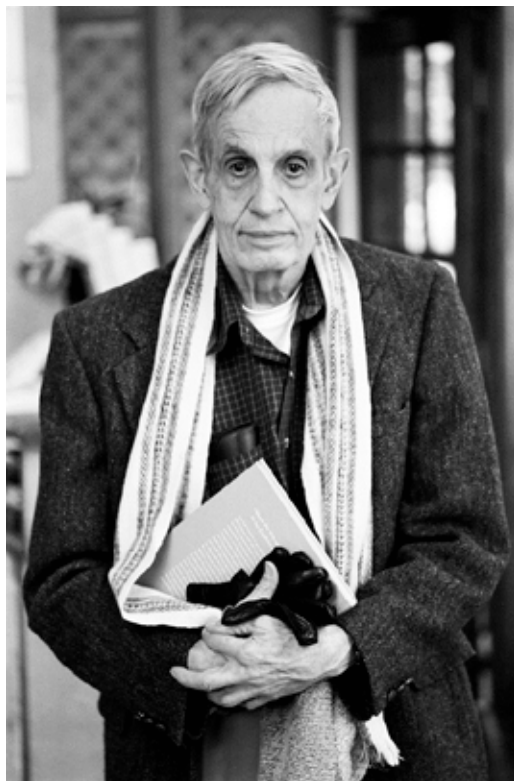


图1 约翰·纳什(1928—2015):非合作博弈论奠基人,诺贝尔经济学奖得主(1994),阿贝尔奖得主(2015)

背。博弈论学家观察到违背纳什均衡的结果应该不会有太多不安,正如任何社会科学专家观察到违背自己认知规律的现象也不会有太多不安(也不一定认为有新的研究机会),因为社会规律都是一种软约束,违背这些规律在逻辑上并非不可能。

所以纳什均衡的实现归根结底还是一种心理游戏。纳什均衡的稳定是一种想象中的稳定。大家都想到了该结果,该结果就真的会实现;大家想不到该结果,该结果也就不会实现。我们知道博弈论的前提假设就是参与人完全理性。我们一般以为理性人是冷冰冰的,而感性人则充满丰富的想象力。其实理性的必然内涵之一就是具有充分的想象力。缺乏想象力的人不可能理性,不可

能只对看得见的事情进行比较和算计就保持基本的理性。人类迄今为止最令人赞叹的想象都来自伟大的理论而非艺术领域。

缺乏想象力势必带来短视。缺乏想象力就无法对未来进行预判，而纳什均衡的解读之一就是“所有人都正确预判了彼此的预判，且没有激励没有能力单方面改变这种局面”。

二、海萨尼转换：想象中的博弈

如果博弈的参与人不仅仅不知道对方的选择，甚至不知道对方的基本信息呢？我们考虑两个人玩石头-剪刀-布博弈。尽管彼此不知道对方究竟要选什么，但是双方以等概率 $1/3$ 选择所有策略是唯一的纳什均衡。如果 A 是一个残疾人永远只会出剪刀，B 是一个正常人但是不知道 A 是一个残疾人，那么这个博弈的均衡结果是什么呢？

在海萨尼之前，类似的问题无解。我们如果站在上帝视角当然会说 B 选择确定性出石头是均衡结果，但是 B 不知道 A 的情况。在海萨尼之前，对这种“不知道”没法建模分析。

海萨尼的做法是通过引入额外可能的博弈把“不知道”建模成一个概率分布。具体做法如下。类似于前面的分析，假设 A 和 B 都是正常人，但是 B 不知道 A 是正常的，那么 A 和 B 之间的石头-剪刀-布博弈依然没法求解：双方均等概率 $1/3$ 选择每个策略并不是均衡结果，因为 B 不了解 A 的情况，没有充分考虑 A 有残疾人的可能。所以在这两个博弈中，尽管 B 只进行了一场博弈，只与一种可能的 A 进行博弈，但是这两个博弈都没法进行孤立分析。

然而两个博弈合在一起就有了分析的可能。海萨尼假设 B 对 A 的两种可能有一个预估（比如有 90% 的可能是正常人，有 10% 的可能是残疾人），通过自然做选择实现 A 的类型，把两个博弈巧妙

地放在一起进行分析，破解了上述难题。

我们知道，海萨尼构建的模型也称贝叶斯博弈，相应的均衡是贝叶斯纳什均衡。引入“类型”和“自然选择”，把信息不完全博弈转成信息完全但是不完美博弈的这个过程就称为海萨尼转换。从信息的视角来看，海萨尼转换是博弈论中的第二个重大突破。

我们初学海萨尼转换理论可能会觉得平淡无奇，因为想象自然在做选择的建模方式在把一些非博弈问题转成博弈问题的时候也经常使用。海萨尼转换如此自然以至于我们目前学习和研究信息不完全博弈或贝叶斯博弈时默认的就是转换后的博弈。

但是对贝叶斯博弈稍微一深入思考以后又会有很多困惑：比如研究前面不完全信息石头-剪刀-布博弈的时候，不仅仅 B 要想象两个可能的对手，A 也要同时想象自己的两种身份：即便明明自己是正常人，也要想象如果是残疾人会怎么办。这其实没有特别奇怪，因为对手 B 在做这种想象，为了分析清楚对手，A 势必也要做这种想象。所以贝叶斯博弈本质是一种想象的博弈，海萨尼转换就是把若干无法孤立求解的博弈通过想象联系在了一起进行统一分析和求解。

提出海萨尼转换无疑需要海萨尼教授的丰富想象力和深刻洞察力，那么贝叶斯博弈中的参与人是否真的像海萨尼教授设想的一样也具有如此丰富的想象力，能把孤立的博弈放到一起思考呢？理论上讲，正如我们前面所提到的，理性人假设已经蕴含了充分想象力。至于精确的先验信息如何获取，即这个例子中参与人 B 对参与人 A 类型分布的预估，则是需要另外解决的问题，此处不做详细讨论。

不完全信息博弈是非合作博弈论的一个重要研究分支，也是信息经济学的基础。正是因为有

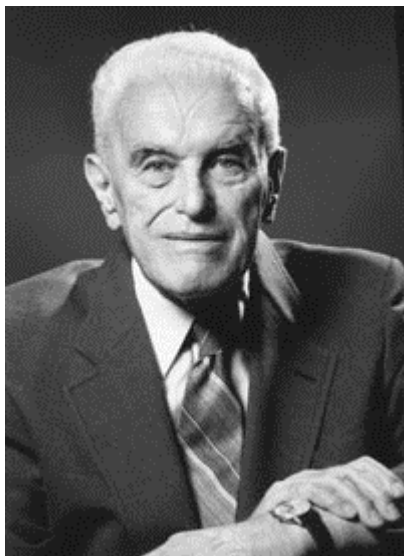


图2 约翰·海萨尼（1920—2000）：著名博弈论学家，诺贝尔经济学奖得主（1994）

了海萨尼转换，我们充分借助理性背后的想象力，把不可能变成了可能。不完全信息博弈以及信息经济学的基本理论今天已经成为数字经济研究中最重要理论工具之一。

三、无知之幕：想象中的公正

海萨尼转换的思想告诉我们，在有些研究中不能站在上帝视角看问题。即便真的拥有所有信息，也只能假设有所不知，只有这样才能真正解决问题。这个道理不仅仅对博弈论研究很有用，也有益于解决很多棘手的社会问题。下面介绍笔者对“无知之幕”这个重要哲学思想的粗浅理解。

无知之幕最初由美国哲学家罗尔斯在他著名的《正义论》中提出，是帮助我们思考如何创造一个公正社会制度的重要思想。笔者不清楚罗尔斯是否熟悉博弈论，但他无知之幕的思想跟海萨尼转换实在联系太紧密了。

公正是整个社会科学探讨的核心议题之一，通常与效率并列为社会科学中最重要的两个议题。

但是公正的概念本身却比效率难理解很多，非常难以形成共识。解决公平性问题大致有两条路径。一条路径是从学理上充分探讨公平公正的内涵，归纳出一些基本的原则，然后基于这些原则解决具体的问题。另一条路径是让不同的利益群体代表各自的利益充分发表各自观点，充分博弈和讨价还价，在一定的框架和机制下解决具体问题。只要相应的过程、框架和机制不先验偏袒任何一方，具有程序正义，那么结果就是公正的。第二条路径其实也离不开第一条路径，因为绕不开何为公正机制的问题。公正机制的判定问题也可以通过投票或者讨价还价的方法来选择，但是选择何种投票机制本身又是争论不清的事情。所以第二条依赖实践的路径也无法摆脱理论探讨。

罗尔斯认为机会平等是公正性的一条基本原则，而无知之幕是论证机会平等的一个重要思想实验。无知之幕要求，探讨公正性问题的时候不要带入自己的任何身份，需要想象成自己躲在一个幕布的后面，幕布拉开后自己有可能是任何一种身份。为保证自己在最差情况下的处境也不



图3 约翰·罗尔斯（1921—2002）：著名伦理学家，政治哲学家，《正义论》的作者

至于太惨，无知之幕背后的提议者会追求机会平等，提出平等主义(egalitarianism)标准下尽可能公平合理的方案。基于无知之幕提出的方案不仅适用于提议者现实中的真实身份，还适用于想象中的所有可能身份；不仅适用于当下，也适用于未来，是一个一揽子的解决方案。

无知之幕的思想跟贝叶斯博弈中任何一种类型的参与人都需要想象自己如果是别的类型会如何做选择具有异曲同工之妙：明明自己拥有明确的身份信息，却要想象成躲在一个幕布背后，有可能“投胎”成任何一种身份。无知之幕的核心思想就是只有无知才能做出公平的选择，不然就很容易有所偏袒。

我们设想高校某学院在讨论年终论文奖励的权重配置问题。有的老师擅长发中文论文，有的老师擅长发英文论文，有的老师则中文和英文同等擅长。我们假设所有人都是风险厌恶的。擅长发中文论文的老师知道自己明年一定发中文论文，提出的奖励方案会倾斜中文论文，而擅长发英文论文的老师知道自己明年一定发英文论文，提出的奖励方案会倾斜英文论文。只有既能发中文论文又能发英文论文的老师，由于明年发表的情况不确定，既有可能是中文也有可能是英文，再加上风险厌恶，提出的方案才更有可能更公平、兼顾中文和英文。

当然，完全有可能学院不存在这种中英文发表比较平衡的老师。但是前面的分析可以告诉我们：把自己置身于一个未来身份不确定的想象中的无知场景会有利于公平分配方案的提出。基于无知之幕逻辑提出的方案更有可能被多数人所接受。

将上述想法推而广之，在一般的游戏规则制定中，只有对自己未来身份信息不确定或者想象成不确定才能迫使方案提出者充分考虑各种情

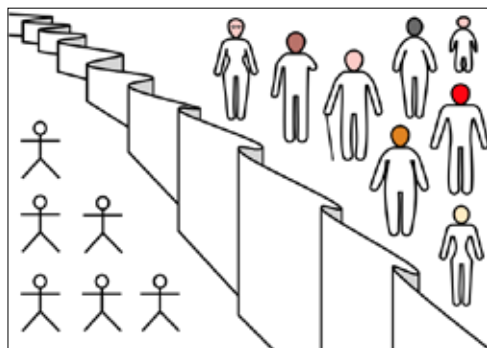


图4 无知之幕示意图

况。这种无知是设计公平公正方案的最大激励。因为你可能是有最多资源的玩家，也可能是资源最少的；你可能是能力最强的，也可能是能力最弱的。为保证自己的最大利益，你会倾向于制定一套尽可能公平的规则，确保即便是最弱势的玩家也不至于输得太惨。如果自己并不是这种情况，未来的身份是确定的，那么基于这个确定身份设计的方案很可能不够公平没法服众，很容易被质疑为自己的利益服务。与此相反，基于想象中的不确定身份而设计的方案则很容易说服别人具有公正性。这就是无知之幕的核心逻辑所在。

四、结语

正如马克思·韦伯所言，“人类是悬挂在自己编织的意义之网上的动物”。人类赋予任何行为和事物以适当的意义和价值，生活的目标就是追求这些意义和价值。人又是一种符号的动物，意义和价值的沟通、传递与交互通常是通过抽象的语言和符号来实现的。充分理解这些意义、价值和符号，无不需要强大的想象能力。有想象才有抽象。

我们从中小学就反复听说想象力很重要，要培养丰富的想象力。但是想象力何以重要，并不

一定所有人都明白。因为对想象力何以重要没有被透彻理解，学术界对科学研究特别是理论研究经常有不恰当的批评：脱离现实、空洞抽象、艰深晦涩，都是经常使用的批评词汇。

笔者当然不是否定有些理论研究的确存在这些问题，只是想强调，这些批评即便符合事实也不必然正当。一定程度的抽象和脱离现实是理论的必然，理论绝不可能只盯住已经发生或者必然发生的历史和现实，理论更重要的是研究所有可能发生的事情。通常只有把历史、现实与可能世界结合在一起才能发现规律、找到解决方案。

理论与抽象的意义何在呢？我们一般认为研究清楚一个具体问题以后，在此基础上进一步抽象提炼出理论。这是科学发现的一个一般过程。本文想提醒读者的是，有时候，没有对抽象问题的理解就没有对具体问题的解决。某些棘手问题的解决需要一个整体视角。很多孤立的局部的现象没有规律，很多孤立的局部的问题也没有解。

正是因为理论研究的是一类问题而非个案，



图5 马克斯·韦伯（1864—1920）：德国著名社会学家，社会学创始人之一

抽象才是必要的。没有抽象的语言就无法叙述这些问题的共性，也就没法深入研究。好的理论一定抽象但是不空洞；相反，里面必然包含了很多丰富的具体内容。好的理论不仅反映现实，更反映可能的世界。即便这个可能的世界永远不实现，事前的整体性研究依然有其必要性。

至于艰深晦涩，这当然是一些理论文章的弊病，但有时候这也只是无法理解抽象语言的批评家的一种托词。对理论论述和抽象语言符号的理解是有门槛的，不仅需要丰富的想象力，还需要清晰的逻辑思维能力和足够的集中精力。

数学是高度抽象的想象的产物：除自然数以外，所有数学都是人类想象的产物。我们以数学为例说明抽象的价值以及多数人对其理解所面临的挑战。把数学改成任何的抽象理论，这种说明大致都还是成立的。由于研究对象的定义需要异常严格，推导过程需要异常严密，数学研究通常使用大量符号和公式。这是多数人眼里数学或应用数学研究的一个显著标志。然而并非总是如此。盖尔和夏普利在促使后者获得2012年度诺贝尔经济学奖的论文College Admissions and the Stability of Marriage中就没有使用什么公式，但是证明一样严谨，没人会否认这是典型的数学论证。

盖尔和夏普利在文章最后讨论了什么是数学思维的问题。他们写到“那么到底什么是数学呢？答案似乎是，只要论证足够严谨，就是数学。你的很多朋友无法理解数学，不是因为他们对数字没感觉，而是因为他们的精力无法集中到足够的程度从而理解一系列的复杂推断。这个观察对于长期教数学的老师并不新鲜，但是此职业以外的人并不一定立即就能接受。对于那些无法接受这一观点的人，本文是一个很好的例证。”（What, then, to raise the old question once more, is mathematics? The answer, it appears, is that any

argument which is carried out with sufficient precision is mathematical, and the reason that your friends and ours cannot understand mathematics is not because they have no head for figures, but because they are unable to achieve the degree of concentration required to follow a moderately involved sequence of inferences. This observation will hardly be news to those engaged in the teaching of mathematics, but it may not be so readily accepted by people outside of the profession. For them the foregoing may serve as a useful illustration.)

在中文里，想象的本意就是“幻想大象”。中国大部分地方并不产大象，但是这并没有阻碍大象这个物种、这个概念在中华文明中所起的作用。想象使我们讨论的有意义的东西变多，抽象则把很多对象归类、很大程度上简化我们的思维。想象离不开抽象，不然我们会陷入细节和具象而迷失方向。抽象更离不开想象，不然理论会陷入空洞而丧失意义。

生命是灰色的，理论之树长青，想象则让生命和理论同时充满了丰富的意义。无知和抽象帮助我们忽略掉各种障眼的细节，后退一步抽身而出，背景退去真相浮出，让这个纷繁芜杂的世界呈现出更加清晰的图景。

致谢：所有图片均来自互联网。

参考文献

- [1] Harsanyi J C. Games with incomplete information played by “Bayesian” players, I–III Part I. The basic model. *Management Science*, 1967, 14(3): 159–182.
- [2] Gale D, Shapley L S. College admissions and the stability of marriage. *The American Mathematical Monthly*, 1962, 69(1): 9–15.
- [3] Myerson, R B. Nash equilibrium and the history of economic theory. *Journal of Economic Literature*, 1999, 37(3): 1067–1082.
- [4] Nash J. Non-cooperative games. *Annals of Mathematics*, 1951, 54(2): 286–295.
- [5] Rawls J. *A Theory of Justice*. Cambridge (Mass.), 1971.
- [6] 刘擎. 西方现代思想讲义. 新星出版社, 2021.



【作者简介】曹志刚，北京交通大学经济管理学院教授。2010年毕业于中国科学院数学与系统科学研究院并留院任助理研究员。2017年9月加盟北京交通大学经济管理学院，任“卓越百人计划”教授。主要研究兴趣为博弈论及其应用，包括网络博弈和算法博弈论等。在相关领域主流刊物发表论文20余篇，包括 *Operations Research*、*Games and Economic Behavior*、*Journal of Mathematical Economics*、*Social Choice and Welfare*、*International Journal of Game Theory* 和 *Theoretical Computer Science* 等期刊以及 *ACM Economics & Computation* 等会议。获中国信息经济学理论贡献奖、系统科学与系统工程青年科技奖、中国决策科学青年科技奖和关肇直青年研究奖。任中国信息经济学会常务理事，中国运筹学会博弈论分会副秘书长、排序论分会副秘书长，中国优选法统筹法与经济数学研究会智能决策与博弈分会副理事长、青年工作委员会副秘书长、网络科学分会副秘书长等。主持包括优秀青年基金在内的国家自然科学基金委课题4项。