

关于博弈论的一些思考： 博弈论的本质及未来

穆义芬 中国科学院数学与系统科学研究院

根据人们的研究，博弈的历史就是生物进化的历史，而博弈决策的历史，就是人类的历史。但博弈论作为一门学科的历史，则只有不足一百年，期间还伴随着许多困惑与质疑。本文希望能回答一些困惑与质疑，能通过对历史的整理回顾来搞明白博弈论的本质、搞清楚博弈论的价值所在，从而找到有待未来研究的真正重要的问题。

本文希望帮助对博弈论有所了解但还有困惑、想要了解和应用博弈论进行交叉研究的人士。

1. 博弈是什么？博弈论是什么？

1.1 博弈论的定义及本质

首先，让我们直接看看博弈论大师们给博弈论下的几个定义。

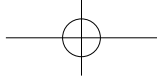
在“Handbook of Game Theory”前言的第一句话中，Robert Aumann和Sergiu Hart开宗明义，写道“Game theory studies the behavior of decision

makers (“players”) whose decisions affect each other”（博弈论研究决策者们的行为，其中决策者们的决策会影响到彼此每一个人）。在《帕尔格雷夫大辞典》中，奥曼对该定义进行了精炼：博弈论是“互动的决策论”。

谢林在《冲突的战略》（P13）写道：我们完全有理由把我们所研究的领域（博弈论）称为“相互依存决策理论”（the theory of interdependent decision）。

罗杰·迈尔森在《博弈论：矛盾冲突分析》中将博弈论定义为“研究智能理性决策者之间的冲突与合作的数学模型。”

所以，当我们谈论“博弈论”时，我们实际上在说什么？我们其实是在讨论“决策”。而从上面的定义中可以看出，博弈的本质，是多人决策，是决策者相互影响、相互依存的决策，是决策者之间存在冲突与合作时的决策。决策者的利益不是完全一致的，他们之间的决策与行为会形



成互相影响的关系，所以任何一个决策者在决策时必须考虑到其他人。

在这里，我们有必要区分人类的两种决策。

第一种是优化决策。这是我们日常生活中每天都要进行的，在进行决策时只有一个优化主体：我们自己。这种一个个体的决策是决策论所研究的内容。

第二种是博弈决策。它是涉及到至少两个人的交互决策问题，因此要求我们必须意识到对方的存在，意识到自己与另外的至少一个人面对面，意识到自己和别人一起身处博弈之中，要共同处理冲突与合作问题。因此，做博弈决策时必须换位思考，“首先要站在对方的角度思考问题”。相比优化决策，博弈决策增加了一个不能抹掉的维度——他人。在这样的场景里，博弈决策是每个决策主体建立决策系统时必须增加的一个模块，并将贯穿在每一个决策环节。

事实上，这种区分正是“Handbook of Game Theory”的前言第一页所描述的：“there is an important distinction between multi-person and one-person decision problems”。与优化决策相比，博弈论在研究对象和适用场景上都具有本质不同。用Aumann和Hart的话说，“this leads to mathematical theories that ultimately yield important and novel insights, quantitative as well as qualitative.”

进一步，博弈论研究的是一个意识到对方也是聪明人的聪明人如何决策的问题。术语里把这种聪明劲儿叫做“理性”。《囚徒的困境》(P59)书中称：“博弈论只研究对赢感兴趣的、有完善的逻辑思维能力的游戏者参与的博弈。只有你相信你的对手（一个或几个）既是理性的、又是希望赢的，而你自己在玩的时候也始终憋着一股劲儿，要力争为自己取得最好的结果，这样的博弈才是博弈论分析的对象。”

博弈论研究的博弈是“不平凡”的。那么，什么是不平凡的？什么是平凡的？用冯·诺依曼的话说：“下棋是‘平凡的’，……下棋不是博弈”（《囚徒的困境》P60）。具体来讲，冯·诺依曼解释道，“象棋不是博弈。象棋是一种定义明确的计算形式。你可能无法算出确切答案，但从理论上来说，一定会有解决的方案，即在任何局势下都存在着一套正确的下法。可是，真正的博弈完全是另一回事。真实的生活也不是象棋中那样的。真实的生活里有虚张声势，有欺诈策略，也会去掂量对方将如何看待自己的意图。这就是我的博弈理论。”（《对赌》P18-19）可见，所谓“不平凡”的博弈，是那些存在策略式行为（欺诈）、因而在其中博弈者必须揣测对方的想法的博弈。

现在，我们想问如下问题，这也是这篇文章最关心的：

在什么情况下博弈论是必不可少、不可替代的？

在这些场景下，博弈论能帮助我们什么？

1.2 博弈论的应用场景及作用

例1：战争、军事、商业竞争等组织/群体之间的对抗/冲突

在有些场景下，典型的如军事/战争，或商业竞争，尽管它们的内在运行细节有非常多的不同，但相同的是一方的博弈结果高度依赖于另一方（的能力和决策）。这时，只考虑自身的优化决策不能帮助我们做出最好的决定，博弈决策思维则是必须引入的，即一定要从博弈的角度来建立决策系统才可能获得更好的或最好的结果。“在战场上，没有博弈意识……甚至会葬送一个国家的未来”（《战略：一部历史》引言）。

那么，在这些场景中，博弈论能帮助我们什么呢？引用《博弈思考法》的话：“清晰的战略

性思考能够帮助你分辨出这些问题或贡献的关键因素。博弈论一个合理的应用就是为解决问题的方案提供思路，并且同时特别提醒人们注意不愿意看到的后果。”（引言）“提高博弈意识同时也能自然而然地丰富你的想象力，从而在今后参与的博弈中规避类似的风险。相比之下，缺乏博弈意识的人更容易陷入战略陷阱，包括其他人有意为他们设下的圈套”（P250，后记）。

例2：囚徒困境或者合作的机制

另一个例子是博弈论中著名的模型——囚徒困境。这一简单的模型是“自博弈论问世以来影响最大也最有争议的一种博弈”。“囚徒困境是一个一般的概念”，它揭示了个体理性与集体理性之间的深刻冲突，“已成为这个时代最基本的哲学和科学课题之一，同人类的生存紧紧地联系在一起。”

表1：囚徒困境的收益矩阵

		Player I	
		C	D
Player II	C	(3, 3)	(0, 5)
	D	(5, 0)	(1, 1)

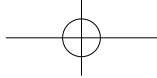
在囚徒困境模型中，由于博弈者是自利的，关心的仅仅是自己的最大利益，因此博弈双方都会选择自己的完全优势策略（dominant strategy）：选择背叛。于是（背叛，背叛）成为纳什均衡。但是，这一结果对博弈者个体和集体来说都比选择（合作，合作）要坏。在这里，个体的理性选择却导致了集体的坏结果，困境由此产生。这种困境具有普遍性，在生活中随处可见。因此，“毫无疑问，囚徒困境模型是所有博弈中被研究最多和被最广泛提及的问题”（《博弈思考法》P23）。《博弈思考法》一书中花了大量的篇幅讨论这一博弈模型，特别是专门有一章节阐述“囚徒困境的重大意义”。另一方面，正

是基于对囚徒困境博弈的深刻理解，人们设计了《反垄断法》中的宽恕法案（Leniency policy），设计了有效的鱼群数量普查机制（《博弈思考法》P166）来限制渔民的捕捞。

另一方面，囚徒困境也是被用来研究人类合作的关键模型。虽然模型预测了困境，但合作却是人类文明的基础。因此，“合作的行为如何演化？(How did cooperative behavior evolve)”这一问题在2005年Science创刊125年时被列为125个重大问题的第16个。对此，密歇根大学教授Robert Axelrod关于重复囚徒困境计算机锦标赛的结果被认为是“对博弈论最有意义的发现之一”。他发现了“一报还一报”策略在重复博弈中是表现最好的。多人的囚徒困境，称为“公地悲剧”，被用于研究公共资源的治理。2009年的诺贝尔经济学奖得主Elinor Ostrom毕生都在研究该问题。她的研究团队对尼泊尔灌溉制度与系统的持续研究说明，即使技术进步和投资增加在短期内能使绩效提高或蛋糕做大，但公共资源使用者之间的博弈依然存在（《共同合作》P95-99）。因此，Elinor声称：“困境永远不会彻底消失，即使在最佳的运作机制中……有效的管理机制比什么都管用。”这表明，博弈——冲突与合作作为人类历史的主题，将一直伴随我们。

1.3 博弈论的核心概念——策略/战略

在博弈中，每一方都调动自己的全部资源和努力来获取最大利益。因此，在博弈的几个要素——博弈者、博弈策略、博弈收益中，策略是最核心的概念。可以说，博弈决策等价于策略决策。也因此，在可行的结构下，虚假信息或者欺骗成为博弈者的必然策略。相应地，机制设计问题被自然地提出以对付这种信息的不对称性。



1.3.1 策略/战略的基本特征

那么，什么是策略？很多文献也将strategy翻译成战略，听起来气势更加宏大。本文将二者通用。根据《战略：一部历史》第三章“起源3：希腊”，“在希腊语中，metis（智慧）描述的是一种特殊的战略智慧概念……希腊语中与之相关的另外两个词语是metiao（考虑、冥想、计划）和metioomai（图谋）。这两个词传递的是一种未雨绸缪、讲究细节、掌握他人思考和行为方式的能力，……同时，它们也包含了欺骗和诡计……”可见，战略或者战略性行为首先包含了有意识的欺骗与反欺骗，这是博弈决策中的一个核心特征。事实上，在冯·诺依曼最初通过扑克建立博弈论时，这一点就作为博弈的核心特征被包含在内。

当然，欺骗具有显著弱点，那就是用几次就不灵了。因此，博弈的不断进行要求博弈者不断变化和扩展自身的策略，这导致复杂系统的出现，而这也是人类的真实历史。在未来的人工智能时代，如何在人机高度融合的环境中，实现战略决策中的欺骗和反欺骗，是一个刚刚萌芽的重要课题。但这一点在目前的博弈论研究中似乎还没有被给予充分重视。

战略的另一个重要的特征是结盟。《战略：一部历史》（P11）中列出了战略性行为所具备的一些基本特征，除了欺骗，还有结盟：实施战略式行为的人，要能“判断对方是潜在的对手还是盟友，”“能通过传递印象或误导等手段来影响他们的行为”，“最有效的战略……还需凭借结盟的能力”。结盟伴随着战争的复杂性而来，“想打赢战争更需要合作与计划”（《战略：一部历史》P38）。但是，结盟具有“作为力量源泉的重要性和不稳定性”，这一点成为了伯罗奔尼撒战争的重要原因（P42-43）。而冯·诺依曼

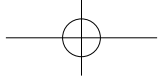
和摩根斯坦在《博弈论与经济行为》中对多人博弈中结盟的研究已经表明，“联盟关系从根本上改变了博弈”，在博弈者数量较多时，结盟将导致“博弈的复杂性与所需的计算量以指数方式增长”（《囚徒的困境》P81）。这带来了联盟博弈研究的极大困难，导致很多年来在博弈论研究中非合作博弈占了主流。

此外，战略还要考虑到博弈中的各种情况，要“关注到所有的可能性”。在其定义里，“策略必须周密地预先描述所有的行动，以致你永远不需要在某一行动之后再做什么决定”。从应用的角度，策略集的完整性等价于博弈模型的完整性，对博弈获胜具有关键作用。在历史案例如四渡赤水中，天才的军事家能洞察到的正是别人看不到的可行策略。反之，如果模型是不完整的，关键的因素被漏掉，则信息优势可能变为决策劣势。因此，在博弈决策系统中一定要考虑到决策系统的鲁棒性，对模型的不完整留有一定的冗余。本文后面，从德州扑克AI程序的研发中可以看出，在设计博弈AI时，通过建立完整的策略模型也是将策略式行为（例如“诈唬”）实现出来的可行方法。

最后，这些战略式行为的基本特征从“早期人类社会和黑猩猩群体”至今，“从来没有发生过改变”，“改变的只是其所应对的局势的复杂性。”（《战略：一部历史》P11）随着时代的发展，对历史中的一些成功策略，我们可能需要扩展其相应的意义。例如远交近攻策略，一般是基于地理距离的；而在新的时代，我们可以在更加抽象的意义上理解和运用，比如在文化或语言上。

例3：边缘战略的历史

《妙趣横生博弈论》第六章中讨论了一种常见的策略——边缘政策（brinkmanship）。这种策



略的核心是在对抗博弈中不断提高威胁：“先发出一个小威胁，然后逐渐提高威胁的规格”。1962年的古巴导弹危机是边缘政策的一个例子；其中的关键在于：“边缘不是险峻的悬崖，而是一道光滑的、越来越陡峭的斜坡”，本质在于“故意创造风险”（《妙趣横生博弈论》P152）。

最早被记载实施的边缘战略出现在《圣经：出埃及记》中（《战略：一部历史》P15-21）。其中上帝通过层层递进的十灾（河水变成血、河里蹦出大批青蛙、虱子灾、苍蝇灾、畜疫灾、泡疮灾、冰雹灾、蝗灾、黑暗之灾、埃及无一家不死人之灾难），给埃及人及法老的心理带来步步深入的影响，从而允许摩西带着以色列奴隶离开埃及。这里，圣经中的上帝在博弈中扮演了聪明而成功的战略家的形象。

1.3.2 信息武器

战略式行为的一个重要表现是对信息的主动利用，这涉及到不完全信息或曰不确定性。“博弈论是本质上不确定的形势中的智能战略”（《战略：一部历史》P152）。“只要当事方关系中涉及冲突、信息不完整，以及有欺骗动机，就和博弈论有关”（《战略：一部历史》P199）。

完全信息博弈研究的是“合理的”博弈结果如何形成，比如囚徒困境博弈中对合作机制的研究。而不完全信息博弈，还要关注信息的处理和利用。这时，除了客观上的不完全信息，还有主观上的不完全信息。对于前者，信息是策略实施的背景；对于后者，信息是策略的一个子集。欺骗与反欺骗，包含着人的能动性，可以理解为“对信息的理解与操纵”，于是信息成为博弈武器。在博弈中，时刻记得你收到的可能都是假信息这一点并不平凡。“兵不厌诈”，这一点在军事中尤其重要。例如，现在假目标仍然是非常有

效的干扰策略。而在导弹上装有诱饵和假目标时如何使用导弹导航系统这一问题如此有趣，竟引起了奥曼对博弈论初始的兴趣。大侦探福尔摩斯在《银额马》《潜水艇计划盗窃案》等案件中也运用了高超的信息甄别与操纵能力。

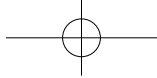
1.3.3 激励或机制设计——对人心的管理

激励或机制设计可以视为博弈的反问题：为了达到想要的博弈均衡结果，应该怎样设计博弈的规则？奥曼将其称为博弈工程学。按奥曼的说法，“一切悲剧都源于不当激励”。每一次社会灾难的背后，都可以看到发生在系统多个环节里的激励不当。在过去的40年，“博弈论与机制设计已成为经济理论的绝对中心”（Eric Maskin），并取得了“博弈论研究中最激动人心的一部分成果”。这些工作说明对博弈系统进行调控时具有博弈思维非常重要。下面我们通过两个简单的例子来说明。

例4：频谱拍卖历史（《频谱拍卖20年》）：拍卖机制设计的复杂性

1990年新西兰首先开始频谱拍卖，采用了Vickrey二级价格密封拍卖机制，即第一报价者得到拍品但只需支付第二高的报价。该机制适用于竞拍者数量足够多的情形，此时诚实报价是竞拍者的最优策略，即拍卖机制是诚实的（truthful）的。但是新西兰拍卖方没有估计到竞拍者不足的情况，也没有设置底价，最终导致了尴尬的结果，如某家公司投标7万多美元，却只支付4美元；某家公司投标500多万美元，但只支付3600美元。

自1990年到2000年，拍卖设计的技术和经验逐渐成熟，在欧洲3G拍卖中，没有再出现频谱贱卖的情况。但英国和德国的拍卖中出现了“赢者的诅咒”现象。具体来说，2000年，竞拍者根据对频谱的估值出价，“价高者得”，但对违约没有设



置惩罚规定，导致两家小运营商以不诚实的高报价获得牌照，而后不断违约，利用规则漏洞反复报出低价。最终政府获得的收入不到违约前报价的一半。

1994年美国进行了频谱拍卖，吸取了新西兰等国家的教训，最终获得70亿美元收入。但是，在1997年的拍卖中，竞拍者利用规则允许报价金额到个位数的漏洞，通过在报价尾数部分携带各自感兴趣的区号向竞拍对手传递瓜分牌照的信息，从而实现了竞拍者间的合谋、避免了价格战，最终导致拍卖收入大幅缩水。

例5：急诊止痛博弈（《博弈思考法》P181-198）

在该博弈中，患者来急诊室，向急诊医生索要某些成瘾性药品作为止痛药。急诊医生则选择要不要给该病人他索要的成瘾性止痛药——可以认为是某种毒品。急诊医生要不要付出巨大的时间成本来判断病人是不是在寻找毒品呢？在判断病人确实是在寻找毒品之后，急诊医生要拒绝他们吗？一方面，拒绝的作用似乎是微小的，因为病人可以去找下一家、下下一家医院。另一方面，急诊医生通常都会受到“病人满意度”评分，他们的薪资甚至工作稳定程度都与满意度得分紧密相关。

因此，主要由于不当激励的原因，急诊医生一般难以拒绝病人的要求而会开出更多的止痛药。这就是典型的“上有政策、下有对策”现象。产生这一现象的根源在于机制设计者没有考虑到医生具有自己的利益，没有把医生与病人之间的博弈考虑进整个系统，而把活生生的人——医生当成了机械系统去调控。这一案例说明，在机制设计中，注意到系统中的每一方都在博弈、都具有自己的利益需求是多么重要的事。

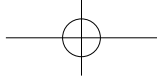
最后，具体环境下的博弈分析和机制设计，

相比于一般博弈理论，特别是一般的激励设计问题，其意义可能更大。这一点与人工智能的研究类似。这一方面是因为激励设计和机制设计问题要研究的是博弈的反问题，因而非常困难：优化问题的反问题已经不是简单问题。另一方面是因为与优化问题不同，博弈论研究的一个很大特点是许多情况下对具体模型的研究更有意义，例如对囚徒困境的研究、对公地悲剧的研究、对拍卖机制的研究等等。正如谢林所说，“在研究过程中，我们不当将问题过于抽象化。如果我们改变博弈场景具体变量的数量，那么我们有可能改变了博弈的特性。因为有些变量可能具有重要价值。”（《冲突的战略》P135）

2. 博弈论的研究方法与博弈决策的困难所在

我们已经看到，博弈决策与优化决策具有本质不同，它们适用于不同的场景，具有不同的研究对象。那么，在研究方法上呢？

博弈决策的首要困难在于意识到对方的存在。你必须时时刻刻记得有一个他人在你的对面，因此这使得第一步的分析中逻辑比传统的数学分析更关键。借用一位经济学教授在博弈论课堂上的话：“game theory is more about logic than mathematics”。这里的mathematics指的是我们熟悉的、数学系学生所学习的那些数学课程。博弈论目前还远远用不到这所有的数学。在《博弈论与经济行为》序言《本书所用方法说明》中，冯·诺依曼称：“所使用的数学方法是初等的……尽管如此，这本书并不真正是初等的，因为本书中的数学推理常常是复杂的，而且广泛地利用了逻辑上的可能情形。……本书所采用的程序的性质，主要属于数理逻辑、集合论和泛函分析”。谢林在《冲突的战略》第六章中声称：



“结果导向的数学结构分析方法不应成为博弈论的主导研究方法。”

此外，博弈需要的“逻辑”是不同于日常生活的逻辑。《战略：一部历史》（P267）中说：

“对立因素之间的相互作用意味着战争领域‘弥漫着一种其自身所特有的矛盾的逻辑（paradoxical logic）’，它与我们生活中其他方面所涉及的一般线性逻辑是对立的。”这使得博弈的分析结果有时候看起来非常简单，但其实看起来显而易见的事情并非那么显然。这一点可以参见拉杆行李箱的发明，谁能想到带轮子的拉杆箱直到1972年才被发明出来呢？

以上这些原因导致博弈论研究面临与优化研究不同的问题或困难。优化已经是成熟的数学分支，而博弈是不能解耦的耦合优化问题。这既导致了博弈决策的复杂度（纳什均衡计算是PPAD-complete），也导致了博弈论研究在其他方面的困难。

另一方面，建立博弈决策系统对人们来说也不是一件“平凡”的事。

首先，策略式思考对人来说并不是天然就能做到的。“人不是天生的战略家。战略需要人有意地做出努力”（《战略：一部历史》P805-806）。按照自由市场的假设，普通人是价格和规则接受方，优化才是自然的决策思维。一般说来，只有受过训练的政治家或军事家才常常保持博弈思维，因为他们的责任天然需要面对冲突或对抗。即使如此，麦肯锡在全球做过调查发现，即使在商界领导人中也有几乎一半人在做出重大商业决策时不会去考虑竞争对手对此如何反应（《博弈思考法》引言）。

其次，人类在漫长的演化过程中形成的认知机制也对形成战略决策系统天然不利。哈佛大学心理学教授丹尼尔·基尔伯特在1991年发现

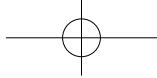
（《对赌》P52）：“来自大量研究文献的发现汇聚在一个观点上，人类是易轻信的生物，他们很容易相信，却很难产生怀疑。”1993年，基尔伯特和同事们通过一系列实验再次证实：“相信我们听到和读到的一切是人类默认的选择。即使一些说法被确切地证实为虚假信息，我们仍然可能将其加工为是真信息”。此外，人类还具有确认偏差，即信念更新会受到既有信念的驱使并不断自我强化（《对赌》P92）。虽然这些信念形成机制是进化的作用，有利于人类生存，但显然这种信念形成过程或者处理信息的方式，是阻碍战略决策因而不利于博弈的。

最后，在实际应用中，更大的问题在于“真实世界中的战略相互作用情况，从不像书中或课堂上给出的典型例子那么简单，你甚至都不清楚自己处于哪一种博弈中”（《博弈思考法》）。这些事实使得博弈论的应用是非平凡的：“如同任何价值极高的东西一样，博弈意识来之不易。它是必须时常培养并保持的一种习惯。”（《博弈思考法》P251）

3. 博弈论与复杂系统

当博弈者通过学习和适应采用策略性行为，带来的必然是动态变化的复杂系统。这一点所带来的万千景象将是超越纳什均衡概念的。

“复杂性理论的最大贡献在于，它……认为个体是复杂系统的一部分，因此评估个体的时候必须与环境结合起来。”（《战略：一部历史》P261）假如我们把“环境”换成“另一个个体”，那么就可以看出：一个博弈系统就形成了一个天然的复杂系统。战略一定是变化的，不然就会被对方利用：“战略从来都不是什么固定的东西，……它是一种包含了重大决策时刻的持续的活动。”（《战略：一部历史》P723）二人或



多人战略的交互演化称为“战略性演变”，被认为是博弈论的第二个焦点，相应的系统是进化的“战略生态系统”（《博弈思考法》P53-63）。

某些条件下，这可能导致系统的均衡，有时会出现状态的周期性行为，如大自然中喉咙颜色不同的蜥蜴所组成的生态系统中出现的进化稳定战略。另外的情形下，系统可能达到“间断均衡”——“相对平静和稳定的状态遭遇暴风骤雨般的重建期”（《战略：一部历史》P722），或者“远离平衡状态”（《金融炼金术》）P30）。这些正是复杂性的真正来源（《系统学是什么》）。也因此，博弈成为导致系统复杂性以及新奇结果涌现的一个核心要素。一个不包含博弈关系的优化系统的运行不会像博弈系统那样导致新的结构。

这进而导致博弈论研究中一个真正的问题：如何建立博弈者的博弈决策系统？这一问题在AI时代显得尤为令人瞩目。

4. AI时代博弈论研究的新机遇

所有的研究，其实都是在试图回答这两个问题：未来的世界是什么样子？实现这一世界的路径是什么样子？现在比较清晰的是，未来的世界，一定是人工智能高度发展的世界，是人与机器融合共处的世界。在这样的世界里，博弈将如何进行呢？人人博弈？人机博弈？机机博弈？都有可能。无论是哪种形式，让机器做出博弈决策，都成为其中的中心问题。

随着人工智能这一波浪潮的兴起，机器博弈已经在各个国家成为被重点关注的研究方向（《2017中国人工智能系列白皮书》）。这当然不仅仅是为了完成各种游戏，而是为了寻求背后的普遍机制和未来更广泛的应用。

技术的变化还带了其他方面的改变。过去，博弈论中的完全理性假设经常被批评，因为现实

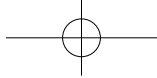
中的人都是有限理性的。也因此，行为经济学和实验经济学得以兴起和繁荣。但是在AI时代，情况可能有所变化。AI具有无限的存储能力、强大的计算能力和疯狂的演化能力，还具有记忆可移植性。那么，一个完全理性的博弈者是不是成为可能了呢？这些基本假设的改变，将给博弈论研究带来一系列的新问题。

4.1 已有的博弈AI及相应算法

游戏AI的发展伴随着人工智能的发展，二者的历史几乎一样长。目前人类已经在很多博弈游戏中实现了人工智能程序（AI），其中最突出的成果当属星际争霸游戏的玩家AlphaStar以及德州扑克的玩家。它们采用了不同的方法。

对于星际争霸这样的游戏，为其设计AI的挑战来自于复杂博弈本身：（1）星际争霸是一个博弈，没有单一的最佳策略；（2）玩家会隐藏信息，必须通过“侦察”积极发现；（3）因果关系不是即时的，早期采取的行动可能无法在很长一段时间内获得回报；（4）玩家必须随着游戏时间的推移不断执行动作；（5）必须实时控制数百个不同的单元和建筑物，从而形成巨大的策略空间。对很多博弈这些要求都是类似的。因此，近年来，星际争霸被认为是最具挑战性的即时战略（RTS）游戏之一，成为人工智能研究的“重大挑战”。

AlphaStar是基于神经网络来构建的，还使用了多智能体学习算法。最初的神经网络由人类游戏的监督学习训练，这样AI可以通过模仿来学习人类使用的基本策略。之后的多智能体强化学习（multi-agent reinforcement learning）过程为：创建一个联盟，联盟中智能体彼此博弈；同时新的竞争者被动态地添加到联盟中，每个智能体都从与其他竞争对手的游戏中学习。为了鼓励联盟中的多样性，每个玩家都有自己的学习目标并在训练



中动态调整。智能体的神经网络权重更新规则则是一种强化学习算法。当然，为了不断探索游戏玩法，AI的战略空间是巨大的。同时，训练过程中AI相当于经历了长达200年的星际争霸实时游戏演化，这显然是人类所不具备的巨大优势。最终的Alpha Star则包含联盟Nash分布的要素，即为已知最优策略的组合。

德州扑克游戏等是不完美信息博弈，其中博弈者不能完全知道对手的策略和博弈局面。在德州扑克这样的博弈中，如冯·诺依曼所说，“诈唬（bluff）是理性的策略”。早在1984年就有职业玩家编写了扑克AI程序，1997年加拿大阿尔伯塔大学的研究人员编写了程序模拟扑克中的诈唬行为，2001年在程序中引入了博弈论方法，2015年引入了反事实遗憾最小化（Counterfactual regret minimization, CFR）算法，解决了两人有限注德州扑克游戏问题。2017年阿尔伯塔大学的DeepStack和CMU的Libratus对算法进行了改进，解决了两人无限注德州扑克游戏问题。2019年CMU和Facebook联手发布了Pluribus，解决了多人不限注游戏问题。这些AI的背后思想其实是比较直观的，可以分成几个部分：第一是对不完美信息动态博弈进行提取/抽象，即对真实博弈进行简化建模，博弈者基于历史状态与信息集进行决策；第二是AI离线自博弈学习，得到初始策略；第三是在线实时搜索改进博弈策略，主要基于CFR算法。CFR算法是一种强化学习算法，给定博弈当前状态、信息集与策略组合，计算博弈者选择某行动与不选择某行动导致的期望收益之差即博弈者的瞬时“遗憾值”（regret），并使其在时空整体上最小。在两人零和情况下，AI走的是近似纳什均衡打法。在多人情况下，AI追踪自己到达当前状态的概率，并通过使用混合策略来实现诈唬。最终，AI能保持自己有一个“极其复杂而

平衡的决策树”，还能发现人类未尝使用的新打法。最近，DeepMind团队在小鼠的大脑实验中发现大脑中也在使用“分布式强化学习”机制。这无疑给AI、特别是强化学习机制的发展应用带来更多希望。

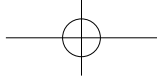
从以上两个博弈AI的实现过程可以看出，博弈中的战略式行为可以通过建立巨大的策略空间和使用混合策略来实现。这给其他更广泛情景下的博弈AI的实现带来了可能。

4.2 博弈AI的一种建立思路

在AI时代，我们需要做的是对每个合适的博弈问题，建立、评估、改进博弈决策系统。这时，相关学科的所有研究成果，如控制论、优化理论、机器学习、因果推断等都可以成为工具。

从博弈AI的算法中可以看出，混合策略对于实现策略式行为起了很大作用。这一想法可以类比人类的“策略基因”演化。人的规则性行为是写到基因里的，就像一个专家系统或数据库。系统/库里面记录了非常多的if情形，以及在各种情形下可行反应then的集合。一个if的string（字符串）要找到对应的then string，每个人的if-then都不同。If string要非常长，以至于能包含在进化中可能遇到的几乎所有境遇。每一个then也是一个string，代表了混合策略。则一个if-then指令实际就是一个（混合）策略，可以称其为策略基因。随着时间发展，if的集合在变，then的集合也在变，一切都在演化。这种建模方式与AlphaStar类似，或许可以将这一想法推广到一般动态博弈系统，作为建立博弈决策系统的基础。

为了建立博弈AI（或博弈决策系统），对环境/局势的评估预测是博弈决策的必要环节。一些游戏AI的成就说明，深度神经网络或许对此有很大帮助。另外，控制论中的各种思想也可能发挥作用。例如，“‘各种对立的特性和力量’相互



交织的时候，智慧是最有价值的手段。基于对现实的充分把握，对未来的明确认识，对往事的经验积累，对不断变化的事物的灵活适应能力，以及对意外事件的充分适应性，智慧适合运用于无法采取既定行动、无法预料行动后果的情况”。

（《战略：一部历史》P37）这种思想和控制论中的PID控制、自适应控制的精髓何其一致。对包含决策系统的动态博弈系统，我们可以考虑构建不同的指标——类似金融市场中的各种指标，来判断或识别系统的涌现特征。

对于模型边界清晰的博弈问题，博弈决策系统可以帮助提高信息处理能力、实现局势演化预测等。但是这里面也蕴含着巨大的风险，比如2019年初的埃航空难事件。此外，2016年3月，在AlphaGo与李世石的第四局比赛中，由于建立AI的过程中必要的剪枝，给AI留下漏洞，导致李世石下出“神之一手”获得了第四局胜利。可以想见，对于具有巨大战略空间的博弈AI，同样的事情一定会再次发生。所以博弈决策系统必须不断更新演化。同时，为保有灵活性和开放性，需要人拥有对决策的最终控制权。相应地，为了实现人与系统/机器的融合，人为此要承受的教育和培训等都要更多。这是AI时代对人类脑力活动的更高要求。

对于边界不清晰的博弈问题，比如人的社交，博弈决策系统或许可以帮助或提醒那些社交障碍患者执行社交对话，并使得“君子可欺之以方”的事情不再发生，从而可以提高社会的公平与效率。

4.3 博弈AI的新问题：技术和伦理

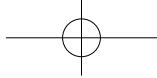
军事和金融是与博弈/战略决策紧密相关的两个领域，也是当今时代和AI时代国家新的国防前沿。在军事中引入AI，建立基于AI的军事智能决策系统，在美国军方已经有很多年的研究了。而

量化交易算法在金融市场中有30多年的历史，并在过去几年迅猛发展，大有取代人类交易员的趋势。

在军事上建立博弈决策系统，一个中心问题是建立人机融合环境下AI的博弈模块，并将其贯穿在决策系统的每一个环节。比如在军事中的OODA循环中，要在Observe（观察）、Orient（调整）、Decide（决策）、Act（行动）每一个阶段中都加入博弈模块，以期实现欺骗与反欺骗功能，从而“保持主动权或者获取主动权”（《战略，一部历史》P259-265）。此外，机器人之间如何实现协作，假脸、假声音技术可能带来什么博弈问题等，都是需要研究的。

AI技术可能导致一些关乎人类整体命运的新问题出现。例如，由于AI能迅速给出对当前状态的反应，于是可能因为反应过快而给现存的系统带来意想不到的影响。比如，军事上的网络战可能在极短的时间内就结束了，这给较弱国家的安全带来了担忧，因而可能引发新的国家间安全协议，新的博弈均衡（《人工智能对国际关系的影响初析》）。再比如金融市场上发生的闪电崩盘（2010, 2013年），（《未来简史》P280-281）：交易算法对消息的反应远远超出人类，当信息发布系统出了故障或被人攻击时，金融产品会在极短的时间内被抛售又被购回。对金融系统的这种冲击是脉冲式的，而人类目前还几乎没有好的应对之策（《系统学是什么》）。

此外，由于AI剥除了人的社会属性，因此AI的价值观或效用是单一的，这导致它很可能是短视的，这不仅可能导致短期利益与长期利益的折中，还会带来诸如自动驾驶的事故责任认定、军事攻击中误伤平民的责任认定等涉及人类社会基本安全的伦理问题，这些都需要人们引入博弈思维才能研究清楚，进而制定相应的伦理规范。



5. 结语

博弈论研究的是博弈决策,这是与优化决策有本质不同的另一种重要决策机制。这使得博弈决策的关键是逻辑推理,采用的是非传统数学分析的方法,并且对人类来说它不是天然就有的思维方式。人类不死,博弈不灭。作为研究智能体间冲突与合作的模型与方法,博弈论将一直伴随着人类。博弈思维/矛盾分析法将一直是人类认识和改变社会的利器。

虽然当前博弈论研究已经大为成功,成为了很多学科的范式和语言,获得了许多次诺贝尔奖,但是无论在理论上还是在应用上对博弈决策的研究都还有很大的探索空间,甚至可以说有一些偏颇甚至错误的认识。为此,我们特别需要澄清博弈决策的基本概念,明晰博弈论的本质与核心。

在AI时代,博弈决策面临新的技术与伦理问题,在重要而适当的问题中建立博弈决策系统是每个组织都要进行的工作。而这正是冯·诺依曼最初创立博弈论时的目标:实现人类理性的可计算化。

最后,博弈论作为一个学科,只是我们认识世界和改造世界的工具之一。为了达到这一目标,必须和已有的一切先进工具与方法进行交叉融合。

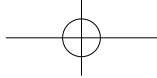
致谢

作者特别感谢赵进钢教授对于本文的总体把关,感谢博弈论学者刘潇、杨扬、王炳昌提出的细致意见,感谢谢亮亮、杨晓光、高小山、王乐一、李俊老师等富有启发性的讨论,感谢张纪峰、洪奕光、刘鲁川、王龙、曹明、黄民懿、吴焦苏等各位师长的鼓励。最后特别感谢恩师郭雷院士,正是郭老师深刻敏锐的洞察力才使得我在

15年前进入到博弈论这一内涵丰富、前景光明的学科。

参考文献

- [1] 劳伦斯·弗里德曼(Lawrence Freedman). 战略:一部历史(Strategy: A History).北京:社会科学文献出版社,王坚、马娟娟译,2016.11.
- [2] Handbook of Game Theory. Edited by R J Aumann and S Hart. Elsevier Science Publishers, B.V. 1992.
- [3] 罗杰·迈尔森(Roger Myerson). 博弈论:矛盾冲突分析(Game Theory: Analysis of Conflict).北京:中国经济出版社,2001.1.
- [4] 戴维·麦克亚当斯(David McAdams). 博弈思考法(Game Changer).北京:中信出版社,杨珮艺、唐源旂译,2016.8.
- [5] 托马斯·谢林(T. C. Schelling). 冲突的战略(The Strategy of Conflict).北京:华夏出版社,赵华等译,2011.5.
- [6] 罗伯特·阿克塞尔罗德(Robert Axelrod). 合作的进化(The evolution of Cooperation).上海:世纪出版集团、上海人民出版社,吴坚忠译,2007.8.
- [7] 阿维纳什·迪克希特(Avinash K. Dixit), 巴里·奈尔伯夫(Barry J. Nalebuff). 妙趣横生博弈论(The Art of Strategy: A Game Theorist's Guide to Success IN Business and Life).北京:机械工业出版社,董志强、王尔山、李文霞译,2012.6.
- [8] 埃莉诺·奥斯特罗姆(Elinor Ostrom). 公共事务的治理之道(Governing the Commons).北京:上海译文出版社,余逊达、陈旭东译,2012.4.
- [9] 艾米·波蒂特(Amy R. Potete), 马克·詹森(Marco A. Janssen), 埃莉诺·奥斯特罗姆(Elinor Ostrom). 共同合作:集体行为、公共资源与实践中的多元方法(Working Together: Collective Action, the Commons, and Multiple Methods in Practice).北京:中国人民大学出版社,路蒙佳译,2011.11.



- [10] 纳西姆·尼古拉斯·塔勒布(Nassim Nicholas Taleb). 非对称风险(Skin in the Game: Hidden Asymmetries in Daily Life). 北京: 中心出版集团, 周洛华译, 2019.1.
- [11] 威廉·庞德斯通(William Poundstone). 囚徒的困境: 冯·诺依曼、博弈论和原子弹之谜(Prisoner's Dilemma), 北京: 中信出版社, 吴鹤龄译, 2015.12.
- [12] 冯·诺依曼, 摩根斯坦. 博弈论与经济行为(Theory of Games and Economic Behavior). 北京: 北京大学出版社, 王建华、顾玮琳译, 2018.5
- [13] 尤瓦尔·赫拉利(Yuval Noah Harari). 未来简史(Homo Deus: A Brief History of Tomorrow). 北京: 中信出版集团, 林俊宏译, 2017.2.
- [14] 安妮·杜克(Annie Duke). 对赌(Thinking in Bets: Making Smarter Decisions When You Don't Have All the Facts). 北京: 中信出版集团, 李光辉译, 2019.1.
- [15] Jingang Zhao, The hybrid solution of an n-person game. Games and Economic Behavior, 1992, 4(1): 145-160.
- [16] 让-雅克·拉丰(Jean-Jcques Laffont), 大卫·马赫迪摩(David Martimort). 激励理论: 委托代理模型(The Theory of Incentives I:The Principal-Agent Model). 北京: 中国人民大学出版社, 陈志俊、李艳、单萍萍译, 2002.6.
- [17] Y.内拉哈里(Y.Narahari). 博弈论与机制设计(Game theory and mechanism design). 北京: 中国人民大学出版社, 曹乾译, 2017.4.
- [18] 乔治·索罗斯(George Soros). 金融炼金术(The Alchemy of Finance). 海口: 海南出版社, 孙忠、侯纯译, 2016.8.
- [19] 傅莹. 人工智能对国际关系的影响初析. 国际政治科学(Quarterly Journal of International Politics), 2019, 4(1): 1-18.
- [20] Oriol Vinyals, Igor Babuschkin, Junyoung Chung, Michael Mathieu, Max Jaderberg, ..., David Silver. AlphaStar: Mastering the Real-Time Strategy Game StarCraft II, <https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii>, 2019-01-24.
- [21] Noam Brown and Tuomas Sandholm. Safe and Nested Subgame Solving for Imperfect-Information Games. In Neural Information Processing Systems (NeurIPS), 2017.
- [22] Noam Brown, Tuomas Sandholm. Solving Imperfect-Information Games via Discounted Regret Minimization, Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2019.
- [23] 马毅华. 频谱拍卖20年: 制度化已形成. 通信世界, 2010, (42): 17.
- [24] Paul R. Milgrom, Putting Auction Theory to Work. Cambridge University Press, Cambridge, 2004
- [25] 郭雷. 系统学是什么. 系统科学与数学, 2016, 36(3): 291-301.
- [26] 2017中国人工智能系列白皮书.
- [27] Dabney W, Kurth-Nelson Z, Uchida N et al. A distributional code for value in dopamine -based reinforcement learning. Nature, 2020, 577: 671 - 675.