

# 基于囚徒困境博弈的 理性、信息与合作分析

刘小山,唐晓嘉

(西南大学 逻辑与智能研究中心,重庆市 400715)

**摘要:** 博弈论为行为哲学的研究提供了有效的理论模型。理性人是博弈论的基础假设,揭示了实现合作共赢上的诸多困境,并引发争论。然而参与者的理性状态同他关于博弈的信息情况相关。通过对囚徒困境这一博弈论经典案例的分析可证明,在特定的信息条件下,合作虽能得到高收益但并非理性博弈参与者的最优选择,理性参与者倾向于不合作;但对博弈进行修正后,参与者的信息情况发生了变化,其最优选择也会发生改变,合作将能够达成并得以保持。声誉、恩仇等机制的引入也提高了博弈达成合作均衡的可能性。

**关键词:** 理性;信息;合作;囚徒困境

**中图分类号:** F224.32 **文献标识码:** A **文章编号:** 1673-9841(2019)01-0021-10

在对社会结构的哲学研究中,如何刻画人的行为动机及其结果是一个非常重要的内容,这关系到对社会结构的哲学研究是否客观与符合实际,研究成果是否具有可操作性以致对制度建设有现实指导意义。博弈论为分析社会中的人如何进行行动策略的选择提供了行之有效的分析模型,提高了相关哲学研究的科学性和实证性。

博弈论有一个基本假设:参与者都是理性的,并且“参与者都是理性的”,这对博弈中的每个参与者来讲都是公共知识<sup>[1]</sup>。这条“理性人假设”带来了许多争论,囚徒困境(Prisoner's dilemma)正是一个典型的案例<sup>[2-3]</sup>。囚徒困境几乎出现在每一本关于博弈论的著作中,它说明了博弈中理性参与者之间的合作在理论上是难以达成的。然而现实中人们在面临囚徒困境时,选择合作的情形却远大于理论上的预计,因此目前囚徒困境博弈研究主要集中在人们为何会达成合作、如何达成合作上,其中分析博弈的重复进行和引入声誉是常见有效的研究方式。

这种研究方式在讨论声誉对囚徒困境的影响时,通常从公共声誉的产生、变更和交易入手,分析其对博弈带来的影响。笔者以此为基础,讨论了一个更特殊的情况,即参与者两两之间的声誉变化对双方的影响,将双方的“私人恩怨”加以刻画,并研究其与公共声誉的区别与联系。由此本文将基于囚徒困境博弈,从理性人假设出发,讨论不同信息情况对博弈参与者理性状态的影响,分析在特定信息情况下参与者合作的可能性。

## 一、理性与信息

博弈论预设了参与者是“理性人”,通常认为其具有以下两个特征:第一,参与者是理性的,即他

收稿日期:2017-01-01

作者简介:刘小山,西南大学逻辑与智能研究中心,博士研究生。

通讯作者:唐晓嘉,西南大学逻辑与智能研究中心,教授,博士生导师。

基金项目:中央高校基本科研业务费专项资金项目“不完全信息重复博弈的动态认知逻辑分析”(SWU1609340),项目负责人:刘小山。

总是选择能使自身效用最大化的行为；第二，每个参与者都知道“每个参与者都是理性的”，并且也知道“每个参与者都知道每个参与者都是理性的”，如此往复下去，也就是说，“所有参与者都是理性的”在博弈中是公共知识<sup>[1]</sup>。

参与者的理性是公共知识这一点，需要在博弈开始之前就得到确认，因为博弈理论必须建立在此基础之上。在理性人假设中，每个参与者都会要求选择能使自己得到最高收益或最小付出的策略，如果某个参与者不是理性的，他的利益要求就不是如此，这样一来他对策略的选择就表现出不确定，由此导致博弈不可能有解。因此，尽管在现实中不理性的情况时常发生，但是在我们的理论分析中，这样的情况会被排除在博弈之外，以保证理论建模的一致性和可行性。

理性人在选择行为的时候只考虑自己能获得的效用，但是据此将理性人与人们平时所说的“自私自利的人”等同是片面的。博弈中的收益一般由效用表示，它是指消费者通过自身行为使自己的需求得到满足的一个度量。人在实际生活中的需求是多种多样的，在追求精神层面上的需求，特别是马斯洛提到的“自我实现需求”时，理性人做出了某个行为，他通过该行为得到了精神上的满足，而另外的人也从他的行为中获益，因此该行为往往会被称作“利他行为”。当一个人通过牺牲自我并帮助他人而获得了精神上的满足感时，这种满足感也可以用效用来反映，一旦该效用（往往是自我价值的实现所带来的满足感）大于他所做出的牺牲（损失的时间、金钱等的效用），那么根据理性人假设，他会很乐意选择做出这样的牺牲；另外，在这个人看来他实际上并没有吃亏，反而是做出了对他来讲最有利的选择。

关于理性，诺贝尔经济学奖获得者罗伯特·约翰·奥曼(Robert John Aumann)有一个经典定义：如果一个参与者在既定的信息下最大化其效用，他就是理性的<sup>[4]</sup>。这意味着参与者的理性状态很大程度由他所掌握的信息决定，参与者根据所掌握的信息确定候选项的效用，并判定对手的策略，在此基础上进行自己最优行动策略的选择。

博弈论所说的信息是指每个参与者具有的与当前博弈相关的知识。博弈中每个参与者的最优选择总是依赖于其他参与者的选择，因此对每个参与者而言，最重要的信息是关于其他参与者的特征和行动的知识。博弈论通常用信息集的概念来对信息进行模型化。一个参与者拥有很多信息集，每个信息集里面有他无法分辨的信息状况<sup>[5]</sup>。例如，一家新干洗店准备在某区域开业，该区域内已有一家旧干洗店。新干洗店不知道自己开业后，旧干洗店是否会跟他打价格战，因此新干洗店在选择是否开业时，“旧干洗店会降价”与“旧干洗店不会降价”同在它的一个信息集中，也就是说新干洗店无法对两者加以区分。另外，旧干洗店知道自己会如何选择，因此在新干洗店开业前，“旧干洗店会降价”与“旧干洗店不会降价”就已分属旧干洗店的不同信息集。

参与者的信息集刻画的是他所拥有的信息，而不是他推断出的结论。例如，新干洗店知道，对于旧干洗店来说，降价的话一定会亏得血本无归，而保持原价的话收益并不会有太大变化，因此新干洗店有充足的理由相信旧干洗店不会降价。但这并不是新干洗店拥有的信息，只是新干洗店根据已有信息做出的对未来的预测，而再合情合理的预测也不等于事实。因此在这种情况下，旧干洗店降价和不降价都在新干洗店的同一个信息集内。

对参与者来说，信息集的子集越多，或每一子信息集内的元素越少，都表示参与者拥有的信息越详细，对博弈的把握也越具体。仍旧以干洗店为例，当新干洗店选择开业后，他关于旧干洗店将要采取的策略的信息集为 $\{\{\text{旧干洗店降价}\}, \{\text{旧干洗店不降价}\}\}$ ，而旧干洗店相应的信息集为 $\{\{\text{选择降价}\}, \{\text{选择不降价}\}\}$ ，可以看到，旧干洗店能够区分自己是否降价的情况，而新干洗店做不到这一点，因此我们可以说，旧干洗店在博弈中拥有更好的信息。

## 二、经典的囚徒困境博弈

囚徒困境是博弈论中的经典案例。以如下模型为例：甲乙两名囚犯被隔离审讯，他们的行动集

均为{背叛(坦白并揭发),合作(抵赖)},用D代表背叛,C代表合作。如果两人都选择策略D,则每人判入狱8年;若只有一个人选择策略D,则他因立功被释放,另一人被判入狱10年;若两人都选择策略C,则两人都将被指控为有罪,但因某些罪名证据不足,每人仅被判入狱1年。在审讯之前,两名囚犯就已经得知了他们行动的所有情况将导致的后果,且两人都明白自己的同伴也拥有与自己相同的信息。如表1所示:

表1 囚徒困境1

囚徒困境		乙	
		C	D
甲	C	-1,-1	-10,0
	D	0,-10	-8,-8

注:二元数组的第一个数表示甲的效用,-1表示被判入狱1年;第二个数是乙的效用

表1中的所有信息都是甲乙两人的公共知识。由于隔离审讯,双方在自己行动前都无法通过观测来了解对方的行动,因此任何一方的所有行动都会包含于对方的同一个信息集之内。这一点也是甲乙两人的公共知识。

可以看出,(C,C)是该博弈的帕累托最优解,因为相应的收益是(-1,-1),对两人来说都是最优的。由此我们会认为(C,C)是对甲乙两人所构

成的集体来讲的最优的策略。但是如果从个体理性出发,这个状态是达不到的。双方在行动前都无法得到对方的行动信息,于是需要根据已有信息对博弈进行推断,而根据理性人假设可以推断出,策略D是每个人的占优策略。因为,当己方选择策略D时,如果对手选择策略C,己方将获得最大收益“0”;如果对手也选择策略D,己方将获得的收益仍相对较高,8年徒刑仍优于10年徒刑。无论对方怎么做,选择策略D都是自己的最好策略。同样的逻辑对另一个人也适用。既然对每个参与者来说策略D始终比策略C更优,那么根据理性人假设可知,双方都会选择策略D,使得博弈的结果为(D,D)。如果使用重复剔除的方式,不管是甲还是乙,都会首先剔除策略C这一劣策略,因此可以得到(D,D)是该博弈的重复剔除优势均衡。

对每个参与者而言,策略D都是对其他参与者策略的最优回应;换句话说,所有参与者都不会对自己的选择感到后悔。选择策略C则不同。当对手选择策略C时,己方若也选择C,收益是“-1”,但选择D则收益为“0”,0要高于-1;如果对手选择策略D,己方若选择策略C,则将面临10年刑期,收益最低。策略C将使每个参与者后悔,即参与者甲和乙都有激励去选择另一策略D以追求更优的收益。然而,与(C,C)相应的收益是(-1,-1),与(D,D)相应的收益却是(-8,-8),个体追求最大利益的理性却导致双方实际得到的收益并非最大,这就是所谓的“困境”。

在一个博弈中如果存在这样的策略组合 $\theta$ , $\theta$ 使得任何参与者在其他参与者不改变已有策略的情况下,都没有激励去改变自身的策略,那么 $\theta$ 就被称为该博弈的纳什均衡。如果一个策略组合是整个博弈的纳什均衡,且它的相关行动规则在每个子博弈上都是纳什均衡,那么它就是该博弈的完美子博弈纳什均衡(subgame perfect Nash equilibrium)<sup>[6]</sup>。显然,(D,D)就是囚徒困境博弈中的纳什均衡。在这样的均衡中,每个博弈参与者都确信,在当前的情况下,他已经选择了最优策略以回应对手的策略。

博弈有完全信息博弈和不完全信息博弈的区分。所谓完全信息是指博弈结构以及每个参与者的收益(收益函数)是公共知识。不完全信息则相反,它是指在博弈中,至少有一个参与者不知道博弈的结构或另一个参与者的收益。博弈的结构和博弈规则的确定可以被看作是自然对博弈进行的设定,因此以上定义可以表述为:在完全信息博弈中,自然要么不首先行动,要么它最初的行动已被所有参与者观察到。在囚徒困境中,甲和乙的信息就都是完全的,表1中的信息全部都是两人的公共知识,这样的信息状态是两人对博弈进行分析的基础。

很多博弈都有一个过程,根据参与者对博弈历史信息把握程度,博弈的信息又分为完美的和不完美的。完美信息博弈是指博弈中的每个信息集都只包含一个结,在这样的博弈中,每个参与者都知道整个博弈的历史,也就是说在博弈中的任何结上,需要选择行动的参与者都知道在这个结之

前博弈进行的整个过程。不完美信息博弈则是指在博弈的某些阶段,要选择行动的参与者并不知道在这一步之前博弈进行的完整过程。

囚徒困境博弈这种存在多个参与者同时行动的博弈是不完美博弈中的典型。如果乙在甲行动之后再选择,并且甲和乙都能观测到对方的行动,那么该博弈就成了完美信息博弈。在这种情况下,甲通过推断可以知道,不管自己做出何种选择,乙都会选择策略 D 以追求最大利益,因此甲会选择策略 D 使自己的利益最大化,使得博弈结果仍然是(D,D)。这样的完美信息对囚徒困境的结果并没有影响,但是一旦博弈中没有重复剔除优势均衡(iterated dominance equilibrium),拥有完美信息将对参与者更有利。

博弈的信息结构还有确定和对称两种分类。在确定信息博弈中,自然不在任何一名参与者行动之后行动,它所带来的不确定性只会出现在所有参与者行动之前。在对称信息博弈中,任何参与者在任何他应该选择行动的终点处,其信息集与其他参与者的相同,也就是说,没有任何参与者在信息上具有优势或劣势。简单囚徒困境中的信息就是确定、对称且完全的,因此在该模型中,得到的所有关于其中一个参与者的结论也同样适用于另一个参与者,只需将结论中甲乙双方双方的所有出现互换即可。

为分析方便,我们将收益符号化,得到上述案例一般描述,如下表 2 所示。

表 2 囚徒困境 2

囚徒困境		乙	
		C	D
甲	C	p, p	q, s
	D	s, q	r, r

注:二元数组的第一个数表示甲的效用,第二个数是乙的效用

该博弈成为囚徒困境需要满足两个条件:(1) $q < r < p < s$  与 (2) $s + q < 2p$ 。在囚徒困境博弈中,参与者的理性使得博弈结果正好为集体利益最低的策略组合,但这并不意味着参与者如果不具有理性就一定能使集体收益提高。

命题 2.1:参与者的不理性可能使集体利益

受损。

可以通过如下方式对命题 2.1 进行简单的证明:当甲乙二人都根据理性行事时,纳什均衡为(D,D),此时的集体收益为  $2r$ 。若某个参与者遵循理性选择了策略 D,另一个参与者因为理性的缺失而选择了策略 C,那么集体收益为  $s + q$ 。由于  $q$  是最差收益,以至无论对条件进行怎样的变形,都不能得出形如  $q > \varphi$  的不等式( $\varphi$  为任意不含  $q$  的不等式)。可以假设  $q$  的取值很小,以至于  $q = r - s$ ,则有  $q < 2r - s$ ,使得  $s + q < 2r$  成立。由此可以看出,参与者的不理性也可以损害集体利益。

如果某个参与者缺乏理性,博弈的过程和结果将变得难以预测。对于保有理性的参与者来说,他仍然可以放心选择那些占优策略,比如在囚徒困境中,他仍然可以放心选择背叛策略 D,因为无论对方做出了怎样的选择,策略 D 都是己方最好的选择。但是理性参与者不能在分析时剔除不理性参与者的劣势策略,因此无法通过剔除各自的劣势策略达到均衡。如果理性的参与者没有优势策略可供选择,也没有劣势策略可供剔除,那么风险规避型的参与者可以考虑回避收益最低的策略,而风险偏好型的参与者可以考虑选择收益最大的策略。

命题 2.2:在不附加限制条件的情况下,囚徒困境参与者无法达成有效的合作型协议。

该命题的证明并不复杂:假设甲乙二人在被捕前订下了攻守同盟协议,即双方都承诺自己的策略只会是 C。以乙在审讯时的选择为例:如果他认为甲会遵守协议选择策略 C,那么对自己而言,遵守协议会得到收益  $\pi(C) = -1$ ,而违背协议可以得到收益  $\pi(D) = 0$ ;由于没有限制条件,违背协议不会对自己的收益有任何附加影响,因此最终  $\pi(D) > \pi(C)$ 。另外,如果乙认为甲会违背协议选择策略 D,那么对自己而言遵守协议会得到收益  $\pi(C) = -10$ ,而违背协议可以得到收益  $\pi(D) = -8$ ,由于没有限制条件,违背协议对收益并没有附加影响,因此同样有  $\pi(D) > \pi(C)$ 。在这样的情况

况下,没有任何激励促使乙遵守协议,因此他会选择策略 D 以追求利益最大化。同样的道理,甲也会选择策略 D,这样一来二人订立的攻守同盟协议就形同虚设。可以看到,囚徒困境博弈中订立没有约束条件的协议是毫无意义的。

命题 2.2 中的协议也可以由单方面的声明来替代,得到如下的推论。

推论 2.1:在不附加限制条件的情况下,囚徒困境中任何参与者的事先声明都不会影响博弈的结果。

该推论的证明与命题 2.2 基本相同,由于理性人假设的存在,因而参与者的事先声明会被其他参与者认为是不可信的。同时,做出过声明的参与者也明白自己的声明并不会被对方接受,因此在博弈时也不会考虑自己的声明被接受的情况。这样一来,参与者的事先声明在博弈中就不会起到任何作用。

命题 2.2 与推论 2.1 有一个重要的前提,“在不附加限制条件的情况下”。在上文的例子中,可以将此理解为甲乙二人在博弈结束之后的人生不会再有任何交集,也不会有任何未参与博弈的人员向博弈的参与者做出与博弈结构无关的声明,如许诺或威胁等。也就是说,囚徒困境中的参与者在做出选择时,不会受到外界或其他参与者在博弈开始前的任何行动的干扰。现实中的情况更为复杂,存在大量单方面许诺或威胁被对方接受的情况,但与其说是该次声明在起作用,不如说是发表声明的参与者的声誉在起作用。未参加博弈的人员发表的声明也可能起作用,那是因为本次博弈的参与者无法确定自己是否会在将来与这些人博弈。关于声誉的分析可参见本文第四部分。

### 三、多次重复的囚徒困境博弈

在现实的社会活动中,存在着大量的人与人之间长期合作的例子。如果囚徒困境博弈重复进行有限次,参与者是否会认识到他们应该合作,于是他们产生某种默契,从而使博弈的结果走向(C,C)呢?对这个问题的讨论可以从博弈的具体次数入手。

#### (一)有限次重复囚徒困境博弈

博弈中的参与者要根据自己掌握的信息规划出自己在未来的策略选择,而且在规划之前要充分考虑其他参与者的规划。有限次重复囚徒困境博弈是具有对称且完全信息的动态博弈,它有以下特征,且这些特征都是参与者的公共知识:a. 没有任何参与者在信息上具有优势或劣势;b. 所有参与者在每次重复博弈中的行动次序与囚徒困境相同;c. 任何参与者在行动之前,已经观测到所有参与者的全部过往行动;d. 所有参与者在每种可能的行动组合下的收益都是公共知识;e. 博弈的次数确定且是公共知识。基于有限次重复的以上特征,求解该博弈时可以使用逆向归纳法(Backward Induction)<sup>[7]</sup>。

命题 3.1:有限次重复囚徒困境博弈有且仅有一个完美子博弈纳什均衡,即所有参与者在每一次博弈中都选择策略 D。

证明:假设囚徒困境博弈重复  $n$  次( $n \geq 2$ )。在第  $n$  次博弈中,由于已经是最后一次博弈,自己的任何行为都不会对以后的博弈有影响,因此参与者的选择与囚徒困境 1 中相同,都会选择背叛策略 D 以追求最优收益,策略组合为(D,D),双方的收益均为  $\pi_n(D,D) = -8$ 。由于双方都知道,无论做什么都不会影响到最后一次博弈,因此在第  $n-1$  次博弈中,双方同样会选择策略 D 以追求最大收益,且不用担心其会对第  $n$  次博弈有任何影响,因此策略组合为(D,D),双方的收益均为  $\pi_{n-1}(D,D) = -8$ 。以此类推,一直到第一次博弈,策略组合都为(D,D),双方的收益均为  $\pi_1(D,D) = -8$ 。这是该博弈的唯一解,也是纳什均衡,同时不管从该重复博弈中取任何子博弈(从第  $k$  次到第  $j$  次,  $1 \leq k \leq j \leq n$ ),其纳什均衡都是策略组合(D,D)的重复。因此,重复  $n$  次策略组合(D,D)是整个博弈的完美子博弈纳什均衡。

当博弈重复进行时,参与者目前的策略选择有可能对以后的博弈产生影响,例如为自己建立声

誉、改变对方参与博弈的积极性等。但有限次重复博弈的一个重要特点在于,参与者在最后一次博弈中无需考虑自己的策略会对以后的博弈有什么影响,因为博弈已经不会再继续重复下去了,这使得最后一次博弈的求解方式与单次博弈相同。在理性的影响下,单次囚徒困境有一个非合作的纳什均衡,因此逆向归纳法总能保证有限次重复囚徒困境博弈有解,且这个解总是一个非合作的完美子博弈纳什均衡。

现实中的“最后一次”通常并不是那么确定,人们在合作愉快之后,说不定会兴致勃勃地商量着再合作一次,而这是双方之前都不确定的。因此可以考虑将有限次囚徒困境的特征  $e$  更改为“博弈有最后一次,且其为公共知识,但每个参与者都不能确定哪一次是最后一次”,在这种情况下,会有以下几种情况出现。

第一种情况,两个参与者都会把每一次博弈都看作最后一次,那么该博弈将变成简单囚徒困境的重复进行,一直重复策略组合(D,D)仍然是整个博弈的完美子博弈纳什均衡。

第二种情况,两个参与者甲和乙分别认为博弈的次数为  $m$  和  $n$ (不妨假设  $m < n$ ),那么由前面的分析可以看出,前  $m$  次博弈时甲都会选择策略 D,若博弈次数超过  $m$ ,甲将重新进行判断;前  $n$  次博弈时乙都会选择策略 D,若博弈次数超过  $n$ ,乙将重新进行判断。如果两人在每次重新判断时都会再选择一个次数,那么策略组合(D,D)将继续重复下去。

第三种情况,其中一个参与者(不妨假设为甲)认为博弈的次数为  $m$ ,另一个参与者(按照假设则应该是乙)不确定到底有多少次博弈。由于乙没有假设博弈次数,他将不能使用逆向归纳法,但前  $m$  次博弈中甲始终会选择策略 D,因此前  $m$  次博弈有一个纳什均衡,即策略组合(D,D)重复  $m$  次。但该均衡对整个博弈来讲是不完整的, $m$  次博弈之后,甲将对博弈次数进行重新判断。如果甲每次重新判断都会选择一个次数,那么他仍然会一直选择策略 D,因此策略组合(D,D)的重复将一直是该博弈的纳什均衡。

第四种情况,博弈双方都不假定博弈的重复次数。这种情况可以理解为,博弈每次重复之后,双方都觉得可能还会有下一次,其分析方式与下文描述的无限次的情况相同。

## (二)无限次重复囚徒困境

如果不管在重复博弈的哪一个阶段,“最后一次”都完全无法确定,那么我们可以对有限重复囚徒困境博弈进行修正,将其次数改为无限次。将模型进行这样处理并不是为了分析真正无限次重复的博弈,而是通过构造一个没有已知结束点的模型,分析现实中那些不确定重复次数,且在每次博弈时,所有参与者都知道这不是最后一轮的博弈。

由于参与者都没有关于最后一次博弈的信息,参与者不清楚自己选择不合作得到的短期收益跟选择合作得到的长期收益相比哪个更优,因此不会有上文提到的那个非合作的纳什均衡,普通的逆向归纳法也无法直接运用。由于心理上的原因和实际利益的影响,人们往往更看重近期利益,而对长期利益相对看轻。但在无限次重复博弈中,只要双方相互配合,长期利益会趋于无穷,因此博弈双方在心理上都会更偏向于获得长期利益。

无限次重复博弈与有限次重复博弈有一个共同点:如果想实现人们理想中的互利互惠的均衡,参与者应当试图“合作”和惩罚“不合作”。在无限次重复囚徒困境中,参与者可以各自采取冷酷策略以得到互相合作的完美均衡(C,C),前提是双方都知道对方在采取冷酷策略。冷酷策略(Grim Strategy)又称触发策略(Trigger Strategy),是指参与者一开始就选择策略 C 并试图重复进行;一旦某个参与者选择了策略 D,其他参与者也将在下一轮博弈中选择策略 D;另外,选择了策略 D 的参与者之后会一直选择策略 D。

命题 3.2:无限重复囚徒困境博弈中,冷酷策略是对冷酷策略的最优反应。

证明:使用表 2 中的一般化数据以使分析具有普遍性。如果甲能够确定乙一开始就会采用冷酷策略,那么对甲来说,第一轮就有两种策略选择,相应的收益分别为  $\pi_{甲}(D, C) = s$  或  $\pi_{甲}$

$(C,C)=p$ 。虽然  $s$  是最优收益,但一旦甲选择策略  $D$ ,乙将从下次博弈开始一直选择策略  $D$ ,甲的收益也将变为最劣  $q$  或次劣  $r$ 。假设甲第一次选择策略  $D$  发生在第  $n$  次博弈,那么在前  $n-1$  次博弈中,每次甲和乙一样都选择策略  $C$  并都得到收益  $p$ ,因此甲在前  $n-1$  次博弈的总收益为  $p(n-1)$ 。又由于甲在第  $n$  次博弈中选择了策略  $D$ ,他将得到由背叛带来的高收益  $s$ ,因此甲在前  $n$  次博弈的总收益为  $s+p(n-1)$ 。从第  $n+1$  次博弈开始,乙将在每次博弈中都选择策略  $D$  且不会再更改,根据理性人假设,甲也会选择策略  $D$ ,得到的收益为  $r$ ,并一直无限重复下去。

令  $x$  和  $y$  是  $n$  的任意两个不同取值,且  $x < y$ ,比较两种取值下甲的收益。当  $n=y$  时,甲在前  $y$  次博弈的总收益  $\Pi(y) = s + p(y-1)$ ;若  $n=x$ ,则相应的可比总收益  $\Pi(x)$  由两部分构成,  $s + p(x-1)$  和  $x$  次后直到  $y$  次的总收益  $r(y-x)$ ,即  $\Pi(x) = s + p(x-1) + r(y-x)$ 。因此,  $\Pi(x) - \Pi(y) = s + p(x-1) + r(y-x) - (s + p(y-1)) = (p-r)(x-y)$ 。因为  $q < r < p < s$  且  $x < y$ ,所以  $(p-r)(x-y) < 0$ ,从而  $\Pi(x) < \Pi(y)$ 。

由  $x$  与  $y$  的任意性可知  $\Pi(n)$  是关于  $n$  的增函数。因此为了使总收益更大,甲要使  $n$  尽可能大,也就是将策略  $D$  的选择拖得越靠后越好。这样一来当  $n$  取正无穷时,甲的总收益达到最大,那么甲在可观测到的有穷次博弈中都选择了策略  $C$ 。以上结果说明,当甲确定乙在一开始就会采取冷酷策略时,没有激励使甲选择策略  $D$ ,因此甲对乙的最优反应是一直选择策略  $C$ ,这正好相当于甲也采取了冷酷策略。

非对称信息在这个博弈中并不一定会损害信息处于劣势一方的利益。假设甲知道乙一开始就会采用冷酷策略,此时甲拥有了更多有用的私人信息,即甲所独有的信息。假设  $\varphi$  = “乙一开始就会采用冷酷策略”,甲的私人信息并不包括  $\varphi$ ,因为乙显然也知道  $\varphi$ 。那些甲知道但乙不知道的信息才是甲的私人信息,唯有那些关于自己和乙的认知情况的信息才是对甲有用的私人信息,如“甲知道自己知道  $\varphi$ ”“甲知道乙不知道甲知道  $\varphi$ ”等。在认知逻辑中通常用  $K_i \varphi$  来表示个体  $i$  知道  $\varphi$ ,这些信息可以形式化为  $K_{\text{甲}} K_{\text{甲}} \varphi$ 、 $K_{\text{甲}} \neg K_{\text{乙}} K_{\text{甲}} \varphi$ <sup>[8]</sup>。甲利用这些信息选择了一个对自己最有利的策略,即自己也采用冷酷策略。显然,相对于甲来讲,乙在信息上处于劣势,但乙并没有损失,反而是获得了更多的利益,因为  $p > r$ 。

#### 四、基于声誉与恩仇机制的囚徒合作博弈

在重复次数有限的案例中,参与者的理性都促使其选择不跟其他参与者进行合作,但并不是说实际情况就和上述分析完全一致,因为这些模型都忽略了实际情况中一个普遍出现的事实,那就是人类社会中对合作的奖励以及对背叛的惩罚。囚徒困境中的参与者之所以难以结盟,原因之一就在于盟约毫无约束力,而参与者更倾向于通过选择背叛来获得较大的眼前利益或规避由对方背叛带来的巨大损失。如果选择背叛的参与者会面临事后惩罚,且惩罚带来的损失大于背叛所产生的利益,或者选择合作的参与者会得到奖励,该奖励带来的收益也大于遭受背叛的损失,那么即使惩罚或奖励要在之后的博弈中才会兑现,盟约也将显得非常有约束力。因为参与者的理性会使他们追求利益最大化,在这种情况下选择彼此合作无疑是明智的。

声誉就是一种常见的约束力,实际上自亚当·斯密以来,声誉一直被认为是保证契约得以顺利执行的重要机制。最早的声誉模型由 Kreps、Milgrom、Roberts 和 Wilson 于 1982 年基于信息的不完全而建立,被称为 KMRW 声誉模型。该模型将参与者分为不同的类型,参与者对其他参与者收益函数或策略空间的不完全信息对均衡结果有重要影响,在这样的情况下,即使是在有限次重复博弈中,只要博弈重复的次数足够多,合作行为仍然得以出现。当声誉机制得以运作时,人们乐于跟拥有好名声的人合作,该合作可最终成功并持续下去<sup>[9]</sup>。

声誉可以被看作一种资产,Kreps(1990)应用重复博弈的无名氏定理对此进行了分析,Tadelis 使用逆向选择(1999)与道德风险(2002)对声誉的交易进行了进一步研究,Mailath 与 Samuelson

(2001)将声誉解释为顾客的预期,并说明了不同能力的企业在交易声誉时有不同的考虑<sup>[10]</sup>。声誉也被看成是反映行为人的历史记录与效用函数的信息,可以降低信息的不对称。

对于重复囚徒困境博弈而言,当次博弈产生的好声誉能在未来带来利益,这样的利益可以看成是本轮策略所产生的收益的延期获取,将这样收益折算在当次博弈中,如果收益足够大,那么合作型策略组合(C,C)将成为纳什均衡。如果声誉信息的传播足够广泛,单次囚徒困境博弈中同样可以达成合作均衡,因为参与者希望在该次博弈中建立的良好声誉能够使他在其他博弈中获益。这解释了现实世界中的大量单次囚徒困境博弈为何会以合作收场。通过放弃眼前利益来获得良好声誉的参与者倾向于促进自身声誉信息的传播,相反,由于占有眼前利益而得到不良声誉的参与者则会倾向于阻止该声誉信息的传播。

在博弈中引入参与者之间的恩仇状况,同样可以使得囚徒困境博弈中关于合作的研究更贴近现实情况。恩仇状况的意思是说,对每个参与者来说,如果对方在之前的博弈中与自己进行了合作(或总体偏向于合作),那么自己感受到对方的好意之后,将更倾向于与其进行合作;如果对方在之前的博弈中背叛了自己(或总体偏向于背叛),自己则会希望对其进行报复。恩仇不同于名声,注重名声的参与者是主动倾向于合作的,而恩仇则表现为参与者根据对手在过去的博弈中的历史行为来确定自己的策略。

在重复囚徒困境博弈中加入参与者相互间的报答与报仇这两样感情因素,令参与者当次选择合作策略C或背叛策略D能得到额外的感情收益,该收益与其他参与者过去的行为相关。

(1)假设乙在过去选择了策略C(这里有两种模型化的方式,即仅指上一次选择了策略C,或者全部的过往选择综合起来看更偏向于合作),此时甲选择策略C得到感情收益 $c_1$ ,选择策略D得到感情收益 $d_1$ ;这里的 $c_1$ 刻画了参与者在与以前的合作伙伴进行合作时的愉快程度, $d_1$ 刻画了参与者在背叛以前的合作伙伴时的愧疚,因此我们根据实际情况设定 $c_1 > d_1$ 。

(2)假设乙在过去选择了策略D,此时甲选择策略C得到感情收益 $c_2$ ,选择策略D得到感情收益 $d_2$ 。这里的 $c_2$ 刻画了参与者在与曾经的叛徒进行合作时的委屈与不安, $d_2$ 刻画了参与者在拒绝与曾经的背叛者合作时的心安理得,因此我们根据实际情况设定 $d_2 > c_2$ 。

(3)比较甲选择策略C时自身的两种不同情况:乙过去选择过策略D或乙过去选择策略C。当乙的过往策略是C时,甲本身就倾向于也选择策略C并得到较高的感情收益 $c_1$ ;当乙的过往策略是D时,甲仍旧选择策略C以试图达成合作,此时甲得到感情收益 $c_2$ ,但是之前遭受的背叛使得这种情况没有前一种情况那么愉快,因此有 $c_1 > c_2$ 。

(4)比较甲选择策略D的两种不同情况:乙过去选择策略C或乙过去选择过策略D。当乙的过往策略是D时,甲本身就倾向于也选择策略D并得到较高的感情收益 $d_2$ ;当乙的过往策略是C时,甲此时背叛了乙,选择策略D以追求当次的高收益,但由于愧疚或心虚等因素,得到的感情收益 $d_1$ 就不高,因此有 $d_2 > d_1$ 。

这样的感情收益由对方的过往策略与参与者自身的当次策略共同决定,但不受对方的当次策略影响。四种感情收益之间差额的大小取决于参与者对恩仇感情的敏感程度。

在加入了恩仇状况的重复囚徒困境博弈中,如果感情收益的影响足够大,那么合作型策略组合(C,C)的重复将成为该重复博弈的纳什均衡。该结论是很容易验证的,当策略组合(C,C)进行重复时,考察博弈中的任意一轮,假设两位参与者处在策略组合(C,C),只要感情收益大到足够保证每位参与者选择策略C的总收益都大于选择策略D的总收益,此时将已选的策略C改为策略D会降低自己的总收益,因此两位参与者都没有激励改变自己的策略选择,那么(C,C)正是该轮博弈的纳什均衡。

该模型并未考虑甲乙双方在过往博弈中结下的恩仇,事实上甲和乙之间的恩仇是可以延续的,并不会局限于某个博弈中。另外,现实中双方的恩仇在博弈中带来的收益往往会随着时间的流逝

而变化(通常是变小)。

与声誉类似,恩仇的引入也是为了促进合作且约束背叛,它可以被近似看作是某博弈中的参与者之间的私有声誉,但恩仇与声誉也有不一样的地方。当甲和乙两人进行博弈时,假设双方有以下公共知识:“乙对丙有恩,乙与丁有仇,甲既不认识丙也不认识丁”。这个时候乙与丙、丁之间的恩仇关系是否会影响甲和乙的博弈?笔者认为不会,因为甲缺少关于丙与丁的信息,不能由他们和乙之间的恩仇关系推断出乙是否值得信赖。而甲在与乙的博弈中建立的声誉,往往会被丙所参考。另外,声誉交易理论将经济主体的声誉看成一种可以建立、投资并管理的资产,并突出了其可交易性<sup>[10]</sup>。参与者之间的恩仇关系并不具备这类性质。

将甲和乙的公共知识换成“乙对丙有恩,丙对甲有恩,甲和乙没有其他关系”,联系现实世界,在这种情况下甲通常会对陌生的乙产生一定程度上的信赖,因此更倾向于与其合作。恩仇状况的这种特点也可以看作是一种只在“恩”这一方面的传递性。而将甲和乙的公共知识换成“乙与丙有仇,丙与甲有仇,甲和乙之间没有其他关系”时,传递关系显然不应该成立,反而应当考虑甲和乙之间是否会因为同仇敌忾而互相产生好感。

更进一步说,将甲和乙的公共知识换成“乙对丙有恩,丙对丁有恩,丁对甲有恩,四人之间没有其他关系”,在这种情况下,由于甲和乙、丙都没有直接关系,而甲和乙之间的关系还是建立在甲与丙的关系之上的,因此我认为在建模时,应当让甲对乙产生的信赖小于甲对丙产生的信赖,也就是说,恩仇状况产生的感情收益应当随着传递而减少。

在以上分析中,甲乙两参与者对于恩仇感情的效用函数是相同的。而在现实中,由于个体在对恩仇感情的敏感程度上存在差异,因而不同的参与者往往拥有不同的恩仇效用函数。假设丙需要在甲和乙之间选择一人进行重复囚徒困境博弈,三者的收益函数是公共知识,收益函数的非感情部分为表2中的数据(由于对称性,三者在这方面的收益是一样的),甲和丙的感情收益分别为  $c_1$ 、 $c_2$ 、 $d_1$  和  $d_2$ ,乙的感情收益分别为  $c_1^*$ 、 $c_2^*$ 、 $d_1^*$  和  $d_2^*$ ,其中  $p+c_1>s+d_1$ ,  $p+c_1^*<s+d_1^*$ 。在这样的情况下,甲与丙能够达成合作型的纳什均衡,即策略组合(C,C)的重复进行;而乙与丙无法达成该均衡,因此对丙来说,与甲进行博弈显然要优于与乙进行博弈。这说明对恩仇感情足够敏感的人能够拥有更多参与合作博弈的机会。在现实中,会碰到参与者对对方的  $p$  与  $s$  之间的差没有准确信息的情况,这个时候,如果有关于对方对恩仇感情敏感程度的信息,也有助于己方判断是否进行博弈。对恩仇感情越敏感的参与者, $c_1-d_1$  的值也越大,越有可能出现  $p+c_1>s+d_1$  的情况,那么达成合作型均衡的可能性也就越高。这也是恩仇状况对博弈参与者的影响因素之一。

## 五、结 语

我们的分析证明,博弈的均衡是在参与者是理性的这一假设与博弈信息状况的双重作用下达成的。根据理性人假设,在最基本的囚徒困境中,博弈的信息状况使合作难以建立;即使过去已经建立了合作,也会因为缺少继续合作的激励而使其难以保持。某个参与者或许会进行事先声明,参与者之间也可能达成协议,但如果这些声明与协议只是希望博弈达成合作,根据理性人假设,这对于参与者来讲都是不可信的。这样的结论与现实情况有一定差距,因此博弈论研究者们构建了重复囚徒困境博弈,希望借此分析博弈双方关于合作的问题。

然而该博弈能否达成合作均衡,还取决于参与者的信息。只有当“博弈无法预测到最后一次”成为参与者们的公共知识,此时参与者们才可能尝试合作。这样的合作仍然不稳定,一旦出现背叛行为,理性的参与者将难以再次达成合作<sup>[11]</sup>。

单纯的重复囚徒困境并不能保证合作,但我们可以对其进行适当的修正,使其更为符合现实情况。声誉的引入为现实中很多看似缺乏理性的合作均衡做出了合理解释。在重复博弈中,某个参与者的声誉可以反映出他是否愿意为了达成合作而放弃短期利益,从而影响对方的选择。引入恩

仇的目的与引入声誉相同,都是为了分析博弈中的合作问题。而与声誉不同的是,参与者对恩仇感情的敏感程度并没有达成统一;且在建模时需要考虑在现实中,人们往往更看重与自己有直接关系或比较近的间接关系的恩仇状况。

关于理性与信息对博弈的影响还有待进一步的研究,如非对称信息博弈、不完全信息博弈、声誉对非理性的影响、匿名交易等方面,都存在很多问题有待解决<sup>[12-14]</sup>。

我们的分析说明,即使存在争议,理性人假设仍然是研究社会个体行为的必要前提,因为它为许多复杂社会现象的研究提供了合理的切入点。在具体问题中,理性人假设刻画了个体行为的动机,以及个体如何理解其他个体行为的动机,且这样的动机以及对动机的理解是可以合理地存在于具体制度之中的。这意味着无论制度设计者的主观愿望如何,一个制度的成功运行是与社会个体利用现有制度机会实现个人合理目标的行为密切相关的。

理性人假设并不是一味地排斥合作,之所以在某些情况下难以达成合作,还因为个体的理性受自身知识状态与信息互动的影 响,它制约着不同信息条件下社会个体的行为选择。在此基础上我们可以认为,社会公平公正的实现不仅依赖于制度的公正与严谨,还依赖于人们在已有信息条件下的理性的行为方式。因此制度设计不仅仅要合理,还需要考虑制度对个体行为方式的影响,以及能否有效地调节社会个人的行为方式。如何在进行制度设计时,让制度既合理又容易被人们接受,让制度在满足个人理性的前提下最大化集体利益,这是社会主义民主制度建设需要思考的一个重要哲学问题。

#### 参考文献:

- [1] VAN BENTHEM J. Rational dynamics and epistemic logic in games[J]. International game theory review, 2007(1):13-45.
- [2] KANAZAWA S, FONTAINE L. Intelligent people defect more in a one-shot prisoner's dilemma game[J]. Journal of neuroscience psychology & economics, 2013(3):201-213.
- [3] 蒋正峰,贺寿南. 博弈论中的理性问题分析[J]. 华南师范大学学报(社会科学版), 2009(1):49-52+68.
- [4] AUMANN R J. War and peace.[J]. Proceedings of the national academy of sciences of the United States of America, 2006(46): 17075-17078.
- [5] 拉斯穆森. 博弈与信息:博弈论概论[M]. 4版. 韩松,张倩伟,庞立永,等译. 北京:中国人民大学出版社,2009:54-56.
- [6] GIBBONS R. Game theory for applied economists[M]. Princeton University Press, 1992:122-126.
- [7] CARROLL J W. The backward induction argument[J]. Theory and decision, 2000(1):61-84.
- [8] 范本特姆. “如实述说”:逻辑中的信息流动[J]. 郭美云,蒋军利,译. 北京大学学报(哲学社会科学版), 2008(1):80-90.
- [9] 李栋,蒋军利,唐晓嘉. 基于名声机制的重复囚徒困境合作博弈分析[J]. 计算机科学, 2013(4):240-243.
- [10] 皮天雷. 国外声誉理论:文献综述、研究展望及对中国的启示[J]. 首都经济贸易大学学报, 2009(3):95-101.
- [11] 李娜,陈安琪. 国外博弈逻辑研究述评[J]. 贵州民族大学学报(哲学社会科学版), 2017(5):122-132.
- [12] KHAN M A, BANERJEE M, RIEKE R. An update logic for information systems[J]. International journal of approximate reasoning, 2014(1):436-456.
- [13] LIU Q. Correlation and common priors in games with incomplete information[J]. Journal of economic theory, 2015(5):49-75.
- [14] ZHANG D, THIELSCHER M. Representing and reasoning about game strategies[J]. Journal of philosophical logic, 2015(2):203-236.

责任编辑 高阿蕊

网 址: <http://xbjbjb.swu.edu.cn>