

# 人类公平感的演化——基于演化博弈论的分析

傅丽虹

广州华商学院数字金融学院, 广东 广州 511300

DOI: 10.61369/SSSD.2025130015

**摘 要 :** 人类是一种天生就具有公平感的物种。本文旨在基于演化博弈论的分析工具, 模拟我们的祖先由原始猿类演化至原始人类的时期, 探讨在不依赖于任何理性假设的前提下, 人类公平感的产生与演化过程。期间在自然演化压力下, 具有一定程度公平感的个体更易繁衍后代, 因此经过2000多万年的演化, 经过突变等动态过程的逐步调整, 公正感的DNA片段具备了演化博弈论上的收敛性, 最终我们祖先都具有这种DNA片段。但在进入了文明时代之后, 随着人类生产力的不断发展, 对人类个体生存繁衍起决定性的选择因素, 逐渐由自然环境转变为社会关系, 人类的自然演化压力已越来越不同于我们的祖先, 因此这种收敛性或许会逐渐消失, 最终该DNA片段会称为被放弃的生物学性状。

**关 键 词 :** 演化博弈论; 公平感; 收敛; 模拟

## The Evolution of Human Sense of Fairness: An Analysis Based on Evolutionary Game Theory

Fu Lihong

School of Digital Finance, Guangzhou Huashang College, Guangzhou, Guangdong 511300

**Abstract :** Humans are an inherently fairness-endowed species. This paper employs analytical tools from evolutionary game theory to simulate the period during which our ancestors evolved from primitive apes to early humans, with the aim of investigating the emergence and evolutionary trajectory of the human sense of fairness—without relying on any rationality assumptions. Under natural selection pressures, individuals possessing a certain degree of fairness had higher reproductive success. Over the course of more than 20 million years, through dynamic processes such as mutation and gradual adaptation, the DNA segment associated with fairness achieved evolutionary stability, ultimately becoming prevalent among our ancestors. However, with the advent of civilization and continuous advances in productive forces, the primary determinants of individual survival and reproduction have shifted from natural environmental factors to social relations. As a result, modern humans face evolutionary pressures markedly different from those of our ancestors. This shift suggests that such evolutionary stability may gradually diminish, and the DNA segment linked to fairness could eventually become an obsolete biological trait.

**Keywords :** evolutionary game theory; sense of fairness; convergence; simulation

### 引言

虽然人类具有后天可塑性, 且很多人在成年后丧失了公平感, 但人类是一种天生就具备公平感的物种。用反证法可证: 假定人类的公平感不是天生的, 那就只能是由人类社会环境所后天塑造的, 由此可推得, 除了人类之外的所有物种都不具备任何程度的公平感。然而实证已证明, 少数的动物也具备一定程度的公平感。故: 虽然人类对公平感的偏好程度的强弱, 很大程度上由后天的社会环境所塑造的; 但是人的公平感的有无, 却是隶属于人的动物性, 是天生的。上述“天生具备”指“公平感存在于人类的DNA中”, 本文旨在用演化博弈论的视角, 探讨为何人类天生就具备公平感。

本论的核心假设是人类公平感可能源于早期人类进化过程中的自然选择, 这种公平感在合作与竞争中具有适应性优势。具体而言, 通过演化博弈论模型和随机演化仿真模型, 探讨公平偏好族群中的演化稳定性; 此外, 筛选压力源自环境条件, 而当原有条件改变后, 演化的收敛方向也随之改变或消失。

## 一、文献综述与理论框架

### （一）文献综述

公平感的形成，需要大量时间形成。在动物中非哺乳类动物皆不具有公平感，只有某几种哺乳类动物具有公平感，少数灵长类动物具备公平感，几乎所有种类的猩猩都具备公平感，而黑猩猩公平感程度高于所有其他动物。据此可合理推论，人类祖先在从普通动物演化为现代人类的过程中，公平感经历了一个漫长形成过程。

资料显示（王胜贤，2023；陈叶烽等，2010），一个物种是具备公平感，取决于该物种的个体间是否具有合作关系；而一个物种的公平感程度的强弱，则取决于合作<sup>[1]</sup>程度强弱。以往关于人类公平感产生原因的讨论（韦倩等，2019；叶航，2015），主要聚焦于文化学，少有将公平感与合作程度相结合。少数用演化博弈论来研究公平感与合作程度的文章，往往直接将人类的合作程度视作一个固定参数（夏巍巍等，2021；罗俊等，2015；韦倩，孙瑞琪，2016）。

与传统博弈论不同，演化博弈论认为参与者有限理性，多数情况下参与者的演化都可以在不依赖于任何理性假设的前提下进行解释（方齐云，郭炳发，2005）。参与者主要通过模仿和突变等动态过程逐步调整策略，以适应环境的变化（王龙，黄锋，2023；董志强，李伟成，2019）。

在由原始哺乳动物演化为现代人类的过程中，我们各代祖先的合作程度，并非是一直不变的，而是经历了一个从无到有、从弱到强的阶段，因此在探讨公平感与合作程度关系时，不应该将祖先的合作程度视作一个外生给定的常量，而应是一个随着时间缓慢增长的内生变量。

### （二）理论框架

本文构建了将合作程度内生化的演化博弈模型，模拟时间跨度为2800至250万年前，在这一过程中，我们的祖先只是不具备思考能力的动物，其决策皆源于本能而非思考。在模拟中，随着时间增加，个体间合作程度不断增大，不同个体在给定规则下做出选择，经多轮优胜劣汰，最终模拟出现代人类对公平感的偏好，是怎样从无到有、由强到弱。

## 二、模型思路与数学形式

### （一）模型思路

首先假定在自然界中，有一种两性繁殖的原始猿类动物，它们出生、捕食、繁衍、死亡。每个个体出生前，它们的两条DNA经历了互换、变异的过程，而它们的表现型（即：策略行为）由DNA唯一决定。在捕食的过程中，它们有时选择自己行动，有时尝试与其他个体合作捕食。在繁衍的过程中，个体所具备的适存度越高，其可能产下的后代就越多，期望的后代数量也就越大。随着时间的变化，一代又一代的个体相继死去。从长期来看，具备更有利于生存的DNA的个体，将产下更多后代；而相对不利于的DNA，将随时间而趋于消亡。

### （二）数学模型

设每代都只能存活N个个体（不妨设N为偶数）。出于讨论方便：设所有个体间都是非血缘且非亲缘关系，且合作只会发生在两个个体间，简化多个个体共同合作的情况。

两个个体间可能会合作捕食，其两个体的总收获以 $\Pi_i$ 表示，也可能独自捕食，其单个个体的收获以 $\Pi_0$ 表示。考虑到在自然界中，单个原始猿类在付出了大量的搜寻成本之后，只能捕获类似于兔子、老鼠之类的小动物，而两个原始猿类合作捕食，却能以相对低的成本，收获类似于野猪、猛犸之类的大型动物，因此假定， $0 < 2\Pi_0 \ll \Pi_1$  且不妨令  $\Pi_1 = 1$ ， $\Pi_0 = 0.01$ 。尽管合作利得显著大于独自捕食，但动物毕竟是一种不懂得思考的东西，它们做出任何决策，都是基于本能选择，因此对于每一个个体，它有时会尝试合作捕食，有事只会单独捕食，所以此模型应存在至少一个混合策略均衡。

当两个个体有意合作捕食时，会涉及到分配问题。假定它们会在合作前，就各自报出自己希望分得份额，若两者份额之和小于等于1，则合作成立，它们将一起捕食；否则合作破裂，它们将选择独自捕食。经过足够长时间演化，最后所形成稳定表现型，即为纳什均衡。用数字  $b_{i,t} (t=0,1,2,\dots; i=1,2,\dots,N)$  代表第t代第i个个体所要求的份额（即该个体的表现型），用向量  $a_{i,t} (t=0,1,2,\dots; i=1,2,\dots,N)$  代表第t代第i个个体的基因型， $a_{i,t}$  为一个  $1 \times 8$  的二进制向量<sup>[2]</sup>。在本文中，假定表现型与基因型完全是一一对应的，其对应关系式为：

$$b_{i,t} = a_{i,t} * (2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7)^T / (2^7 - 1)$$

因此， $b_{i,t} \in [0,1]$ 。且  $\forall i \neq j; i,j=1,2,\dots,N$ ，有：

$$\begin{cases} b_i + b_j \leq 1, & \text{则合作成立} \\ \text{若 } b_i + b_j > 1, & \text{则不合作} \\ b_i + b_j < 1, & \text{则 } 1 - b_i - b_j \text{ 部分被浪费掉} \end{cases}$$

为后续讨论方便，将  $b=1$  个体称为“完全自私的个体”；将  $b < 1$  个体称为“具有公平感的个体”；将  $b=0.5$  个体称为“绝对公平的个体”；将  $b < 0.5$  个体称为“无私的个体”。

人类的演化是一个很复杂的过程，但简单来说，模型可设定为：假定一个个体一生中捕食M次，而  $m (m=1,2,\dots,M)$  为一个个体一生中在捕食前尝试寻找合作伙伴的次数，设：

$$m = m(t), \frac{\partial m}{\partial t} > 0, m \text{ 为非负整数}$$

不妨令  $m = \lfloor t/100 \rfloor$ 。则第t代第i个个体一生总收益为：

$$\Pi_{i,t} = (M - z_{i,t}) * \Pi_0 + z_{i,t} * \Pi_1 * b_{i,t}$$

$z$  为一生中实际合作次数，是内生变数，其值具有一定程度随机性，且  $z \in [0, m]$ 。

在第t期末，个体开始繁衍后代。第t+1期任一个体j，有  $P_{i,t}$  概率继承第t代第i个个体的基因型， $i,j=1,2,\dots,N$ ，且：

$$P_{i,t} = \frac{\Pi_{i,t}}{\sum_{n=1}^N \Pi_{n,t}}$$

令  $c_{i,t+1}$  为第t+1代第i个个体所继承基因， $c_{i,t+1}$  由上述过程生成，因此  $c_{i,t+1}$  也是一个  $1 \times 8$  二进制向量。当N个  $c_{i,t+1}$  向量都生成完毕后，这N个向量会两两随机配对一次，且在配对后， $c_{i,t+1}$  上

的每个元素都以  $x \in [0,1]$  的概率与对方同位置的元素互换<sup>[3]</sup>。配对结束后，每个元素再以  $y \in [0,1]$  的概率产生突变<sup>[4]</sup>。在互换与突变都结束后，第  $t+1$  代的个体  $i$  则得到期最终的基因型，表示为  $a_{i,t+1}$ 。

### 三、电脑模拟模型的设定与结果

以电脑模拟为工具，证明公平感产生有可能只是种间合作演化产物；且当种间合作压力消失后，人类公平感趋同性也会随着演化进程而消失。

#### （一）命题一的设定与结果

命题一：随着时间的流逝，若物种内个体间的合作程度逐渐增加，则一个初始时由完全自私的个体组成的物种，只要演化的时间足够久，最后会演化成一个几乎以完全公平的个体为主体的物种。

##### （1）命题一的设定

设定  $N=1000$ ,  $M=100$ ,  $x=0.5$ ,  $y=0.01$ ,  $t=10000$ 。

##### （2）命题一的结果

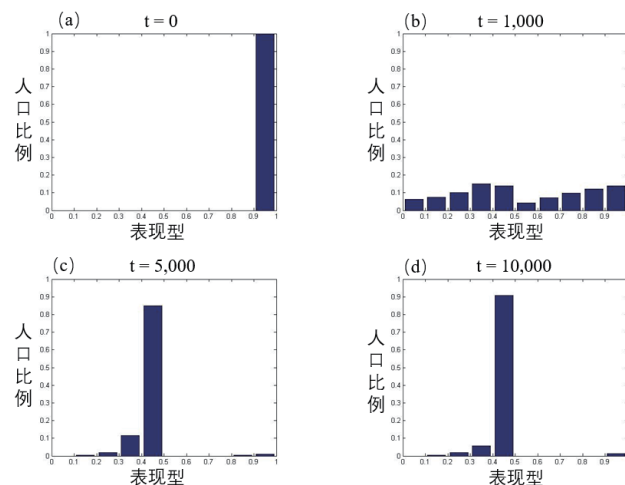


图1 命题一的电脑模拟结果

由图1可见，在  $t=0$  到  $t=10000$  的演化过程中：当  $t=0$  时，所有表现型皆为  $b=1$ ；当  $t=1000$  时，已产生各种不同的表现型；当  $t=5000$  时，已有明显的向 0.5 收敛的趋势；当  $t=10000$  时，收敛趋势更加明显。

由此可证，随着时间的流逝，若物种内个体间的合作程度逐渐增加，则一个初始时由完全自私的个体组成的物种，只要演化的时间足够久，最后会演化成一个几乎以完全公平的个体为主体的物种。

#### （二）命题二的设定与结果

命题二：假设随着时间的继续流逝，人类科技的日益发达，使得个体间的合作需求逐渐减少至消失。当合作与否皆不产生演化压力时，人类公平感的趋同性也会随着演化进程而消失，只要时间足够长，每个人类的公平感程度是随机的。

##### （1）命题二的设定

设定  $N=1000$ ,  $M=100$ ,  $x=0.5$ ,  $y=0.01$ ,  $t=10000$ 。已知  $t=0$  时  $m=100$ ，假定  $m$  每一期减少 1，直至  $t=100$  时  $m=0$ 。

#### （2）命题二的结果

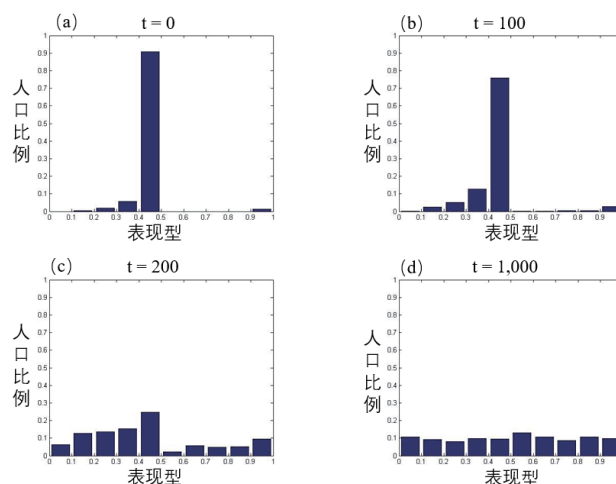


图2 命题二的电脑模拟结果

由图2可见，在  $t=0$  到  $t=1000$  的演化过程中：当  $t=0$  时，所有表现型趋于 0.5；当  $t=100$  时，表现型的集收敛程度已经减少，但仍具有收敛性；当  $t=200$  时，已有明显的随机性；当  $t=1000$  时，随机性更加明显。

由此可证，假设随着时间的继续流逝，若个体间的合作需求逐渐减少直至消失，则人类公平感的趋同性也会随着演化进程而逐渐消失，且只要时间足够长，每个人类天生的公平感程度是随机的。

### 四、结论

本文基于演化博弈论，通过电脑模拟的方式证明，即使不依赖于任何理性假设，在仅存在继承与突变的前提下，一种不具备公平感的物种，随着时间的流逝，若在自然选择的压力下，物种内个体间的合作程度逐渐增加，则只要演化的时间足够久，最后会演化成一个几乎以完全公平的个体为主体的物种。从演化博弈论的视角来看，这恐怕也是人类公平感的产生过程，即一种在长期博弈下的收敛过程。

然而，在我们进入文明时代之后，随着科技的不断发展，人类的自然演化压力已越来越不同于我们的祖先，那种“不合作捕食就有饿死的危险”的自然选择压力早已逐渐淡化，乃至消亡。因此，从演化的角度来看，公平感的重要性已远不如从前，这种收敛性或许会逐渐消失，最终该 DNA 片段会成为被放弃的生物学性状。而我们的电脑模拟结果，也证实当个体间的合作需求逐渐减少直至消失，则人类公平感的趋同性也会随着演化进程而逐渐消失，且只要时间足够长，每个人类的公平感程度是随机的。

#### 附录一

对于“一个  $1 \times 8$  的二进制向量”的举例说明：

向量  $(0,1,1,1,1,0,0,1)$ 、 $(0,1,0,1,1,1,0,1)$  等，都属于  $1 \times 8$  的二进制向量；

向量  $(0,1,2,1,1,0,4,1)$ 、 $(0,1,0,1,8,1,0,1)$  等，不属于

1×8的二进制向量；  
向量 (0,1,0)、(0,1,0,1)、(0,1,0,1,1,1) 等，也不属于  
1×8的二进制向量。

附录二

对于“互换”的举例说明：

两个基因组  $\begin{cases} (0,1,1,1,0,0,1) \\ (0,1,0,1,1,0,1) \end{cases}$ ，可能会互换成为：

$$\begin{cases} (0,1,1,1,1,0,1) \\ (0,1,0,1,1,0,0,1) \end{cases}$$

也可能互换成为：

$$\begin{cases} (0,1,0,1,1,0,0,1) \\ (0,1,1,1,1,0,1) \end{cases}$$

但不能互换成为：

$$\begin{cases} (1,1,1,1,0,0,1) \\ (0,1,0,0,1,1,0,1) \end{cases}$$

附录三

对于“突变”的举例说明：

基因组 (0,1,1,1,1,0,0,1)，可能会突变为  
(0,0,1,1,1,0,0,1)，也可能会突变为 (1,1,1,1,1,0,1)，甚至  
在概率极端小的情况下可以突变为 (1,0,0,0,0,1,1,0)，但无  
论怎么突变，基因组都仍然是一个1×8的二进制向量，因此不能  
突变为 (0,1,1,1,2,0,0,1) 或 (0,1,1,1,1,0,0)。

参考文献

[1] 陈叶烽,叶航,汪丁丁.信任水平的测度及其对合作的影响——来自一组实验微观数据的证据[J].管理世界,2010,(04):54-64.  
[2] 董志强,李伟成.禀赋效应和自然产权的演化:一个主体基模型[J].经济研究,2019,54(01):182-198.  
[3] 方齐云,郭炳发.演化博弈理论发展动态[J].经济学动态,2005,(02):70-72.  
[4] 罗俊,叶航,汪丁丁.捐赠动机、影响因素和激励机制:理论、实验与脑科学综述[J].世界经济,2015,38(07):165-192.  
[5] 王龙,黄锋.多智能体博弈、学习与控制[J].自动化学报,2023,49(03):580-613.  
[6] 王胜贤.群体合作演化的激励控制与优化[D].电子科技大学,2023.DOI:10.27005/d.cnki.gdzku.2023.000074.  
[7] 韦倩,孙瑞琪.强互惠理论研究前沿动态[J].山东行政学院学报,2016,(03):77-83.  
[8] 韦倩,孙瑞琪,姜树广,等.协调性惩罚与人类合作的演化[J].经济研究,2019,54(07):174-187.  
[9] 夏巍巍,罗俊,叶航,等.宗教信仰与信任:一个实验经济学的研究[J].中国经济问题,2021,(01):125-141.  
[10] 叶航.利他行为的经济学解释[J].经济学家,2005,(03):22-29.