

# 收益导向博弈：长期收益最大化的动态框架

黎工

## 摘要

传统博弈论包含两大主要分支：合作博弈与非合作博弈。二者均预设行为模式——要么合作，要么非合作。本文提出**收益导向博弈 (POG)**，即以长期收益为核心的分析框架，不施加任何事前行为假设，并将行为视为动态调整的手段。本文定义四个核心变量：贡献  $C$ 、诉求  $D$ 、议价能力  $W$ 、随机冲击  $\xi$ （区别于现有文献中的认知偏差）。研究证明：在长期互动中存在唯一最优诉求  $D^* \in (0, 2)$ ，介于完全合作 ( $D = 0$ ) 与完全对抗 ( $D = 2$ ) 之间。该点的存在性可通过数学证明，但其取值通过动态调整涌现，而非单纯静态最优化。

框架揭示一项基本结论：个体在追求自身长期利益时，会自然表现出看似利他的行为，例如对特定对手表现出慷慨。这种“自利导致利他”并非源于道德规范，而是内生互动结构的产物。本框架将博弈论核心问题从“均衡点在何处”转向“参与者如何动态行动”，更贴合现实策略行为。

**关键词：**收益导向博弈；动态框架；长期收益；诉求；对手选择性；自利致利他；纳什均衡

## 1 引言

现实中，行为主体极少面临纯粹合作与背叛的二元选择。在长期关系中，个体实际行为是根据反馈动态调整行为：观察对方反应、评估自身得失，继而决定下一步行动。个体基于反馈与收益调整行为，最终收敛至双方均可接受的状态。

自冯·诺依曼与摩根斯顿、纳什以来，现代博弈论沿两大范式发展。合作博弈假设个体可达成有约束力协议，关注联盟形成后的收益分配。非合作博弈则假设个体独立行动，在给定他人策略时寻找无人愿意单方面偏离的点，即纳什均衡。

两类范式具有共同特征：均事前预设参与者行为——一为合作，一为非合作。这与现实观察不符。现实中个体动态试探、适应反馈，并通过反复互动形成可接受结果。行为经济学与演化博弈长期强调此类适应性动态，但以长期收益为核心的统一框架仍未建立。

本文填补这一空白，提出**收益导向博弈 (POG)**：以长期收益最大化为目标，不施加任何事前行为假设，并将行为视为连续可调工具。定义四个核心变量：贡献  $C$ 、诉求  $D$ 、议价能力  $W$ 、随机冲击  $\xi$ （代表环境不确定性，区别于现有文献中的认知偏差）。本文证明：长期互动中

存在唯一最优诉求  $D^* \in (0, 2)$ ，位于完全合作与完全对抗之间。该点存在性可数学证明，但其具体取值通过长期动态调整涌现。

分析得出更深层结论：特定条件下，最优策略可表现为慷慨，即一类外观上的利他行为。这种“自利导致利他”并非外部道德强加，而是互动结构内生出现，尤其来自对手选择性参数  $\lambda$ 。该结论为纯自利动机下合作行为的涌现提供微观基础。

## 2 核心公式

$$\text{收益} = F(C, D, W, \xi)$$

其中：

- 收益：长期总收益（目标函数）
- $C$ ：贡献，满足  $C \in [0, 1]$
- $D$ ：诉求，满足  $D \in [0, 2]$
- $W$ ：议价能力
- $\xi$ ：随机冲击（环境不确定性， $[\xi] = 0$ ）

长期收益取决于个体投入  $C$ 、诉求水平  $D$ 、议价能力  $W$  与随机冲击  $\xi$ 。参与者唯一可主动调整的变量为  $D$ ，其余为外生或随机变量。

## 3 变量定义

### 3.1 贡献 $C$

贡献  $C$  为参与者在博弈中投入的资源、努力或投入比例，满足  $C \in [0, 1]$ 。 $C = 0$  表示对结果无任何贡献， $C = 1$  表示对联合结果完全贡献。

### 3.2 诉求 $D$

诉求  $D$  定义为个体要求回报与自身贡献之比：

$$D = C$$

其中 为个体要求分得的总剩余  $= C + C$  份额。对称贡献下 ( $C = C$ )， $= 2C$ ，故  $D \in [0, 2]$ 。

- $D = 0$ ：不要求回报（完全合作）
- $D = 1$ ：诉求等于贡献（公平交换）
- $D = 2$ ：独占全部收益（完全对抗）

$D$  是参与者**唯一可主动调整**的变量，也是本文核心决策变量。分配中，双方  $D$  共同决定单期收益份额。本质上， $D$  为分配权重，数值上等于要求回报与自身贡献之比。

### 3.3 议价能力 $W$

议价能力  $W$  代表个体在谈判中的影响力，决定诉求在分配中实现的程度，满足  $W \in [0, 1]$ 。

- $W = 0$ ：个体完全无分配权
- $W = 1$ ：分配完全按照个体诉求  $D$  执行

### 3.4 随机冲击 $\xi$

随机冲击  $\xi$  表示外生环境不确定性、随机事件等不可控因素，满足  $E[\xi] = 0$ 。注意  $\xi$  代表外生不确定性，区别于相关文献中的内生认知偏差。

## 4 模型假设

本文聚焦参与者**唯一可主动调整**的变量：诉求  $D$ 。构建数学模型时，其余三个核心变量处理如下。

### 4.1 贡献 $C$

贡献  $C$  决定个体初始位置，并通过  $\pi_i = C_i + C_j$  影响单次互动总剩余。 $C$  对收益的影响是乘性的： $C$  变化使  $\pi_i$  成比例变化，而极值点  $D^*$  位置由  $(D)$  与  $(D)$  互动决定。 $C$  的缩放效应体现在  $(D)$  中。因此， $C$  具体取值不影响“唯一最优解存在”这一定性结论，仅改变最优解  $D^*$  的数值位置。

**假设：**设  $C = C = 0.5$ ，合作总剩余  $\pi = C + C = 1$ 。在确定  $D$  取值区间时，将  $C$  视为常数具有理论合理性，因其变化仅平移  $D^*$ ，不改变其存在性。

### 4.2 议价能力 $W$

议价能力  $W$  影响分配中的权重。在对称情形下 ( $W = W = 0.5$ )， $W$  在分配规则  $\pi_i = D_i D_j + D_j D_i$  中被约去，不进入收益函数。若  $W$  非对称，则需要新的分配规则，会改变  $(D)$  具体形式，但不改变单期收益随  $D$  递增的基本性质。因此， $W$  具体取值同样仅影响  $D^*$  位置，不改变结论定性特征。

**假设：**设  $W = W = 0.5$ 。在区间分析中将  $W$  视为常数是合理的，因为对称假设下  $W$  不参与收益计算，不影响  $D^*$  推导。

### 4.3 随机冲击 $\xi$

随机冲击  $\xi$  以加法形式进入收益函数：

$$\text{收益} = \text{确定性部分} + \xi, [\xi] = 0$$

在求解最优诉求区间时，随机扰动期望为零，不改变极值点位置。因此，在确定性优化中忽略  $\xi$  是合理的；其仅影响收敛路径波动，不改变  $D^*$  本身取值。相关拓展留待未来研究。

## 5 收益结构：从单次互动到长期

长期收益并非单阶段收益简单加总，而是由单次互动收益与持续互动概率共同决定。

### 5.1 简化长期收益函数

在第4节模型假设下，长期收益函数可简化为决策变量  $D$  的单变量函数：

$$\text{收益} = (D) = (D) (D)$$

其中：

- $(D)$ ：单次互动收益
- $(D)$ ：持续互动概率
- $:$  互动次数

### 5.2 单次互动收益 $(D)$

设双方诉求分别为  $D_1$ 、 $D_2$ ，合作总剩余  $= 1$ （由  $C_1 = C_2 = 0.5$  且  $= C_1 + C_2$  得出）。单次互动收益按双方诉求比例分配：

$$(D) = D_1 D_2 + D_2$$

注意  $D_1$ 、 $D_2$  是决定分配比例的权重，而非要求总量。该分配规则类似纳什议价中的加权比例解， $D$  反映个体议价能力。

给定对方诉求固定为  $D_{\text{oter}}$ ，个体单次互动收益为：

$$(D) = D D_{\text{oter}} + D_{\text{oter}}$$

该函数满足：

- $(0) = 0$
- $(D)$  关于  $D$  严格递增：

$$(D) = D_{\text{oter}}(D + D_{\text{oter}})^2 \geq 0$$

### 5.3 持续互动概率 $(D)$

持续互动概率  $(D)$  定义为诉求  $D$  的函数，满足基本边界条件：

- $(0) = 1$ （完全合作保证关系持续）
- $(2) = 0$ （完全对抗导致关系终止）

引入对手选择性参数  $\lambda$ ：  $\lambda = 1$  为线性容忍；  $\lambda < 1$  表示相对宽容；  $\lambda > 1$  表示高选择性，小幅偏离公平即终止合作。

本文采用如下函数形式：

$$(D) = \left(1 - D2\right)^{\lambda}, D \in [0, 2]$$

该函数满足：

- $(0) = 1$
- $(2) = 0$
- $(D) = -\lambda 2(1 - D2)^{\lambda-1} < 0$ （严格递减）

$(D)$  可视为简化形式的接受概率。其具体形式可由微观基础导出。为便于处理，本文采用幂函数形式以获得闭式解。

## 5.4 长期收益公式

长期收益为单次互动收益、持续互动概率与互动次数 的乘积：

$$\text{收益} = (D) (D)$$

代入  $(D)$  与  $(D)$  具体形式：

$$\text{收益} = DD + D_{\text{oter}} \left(1 - D2\right)^{\lambda}$$

将 视为常数，可吸收为比例系数，不影响  $D$  最优化。因此最优化问题简化为：

$$(D) = DD + D_{\text{oter}} \left(1 - D2\right)^{\lambda}, D \in [0, 2]$$

其中  $D_{\text{oter}}$  为对方诉求，本文视为固定参数  $D$ 。

## 5.5 动态学习与策略演化

在上述假设下，有限理性个体通过长期互动逐步调整诉求。设  $D()$  为第 期诉求，学习规则为：

$$D(+1) = D() + (D())$$

其中  $\in (0, 1)$  为学习率。为保证收敛，需满足  $2 D \in [0, 2] |(D)|$ 。该规则表明：

- 边际收益为正，个体提高诉求
- 边际收益为负，个体降低诉求
- $(D()) = 0$  时个体停止调整，达到个体最优

多主体同时更新时，该系统属于超模博弈类。省略正式收敛结果。

## 6 核心结论数学证明

### 6.1 收益函数

基于第5节，长期收益函数为：

$$(D) = DD + \left(1 - D2\right)^{\lambda}, D \in [0, 2]$$

其中  $= D_{\text{oter}}$  0 为对方诉求， $\lambda$  0 为对手选择性参数。最优化问题为：

$$D \in [0, 2] (D)$$

### 6.2 一阶导数

为简化符号，记  $= D +$ ，则：

$$(D) = D \left(1 - D2\right)^{\lambda}$$

对  $D$  求导：

$$(D) = {}^2 \left(1 - D2\right)^{\lambda} - D \lambda 2 \left(1 - D2\right)^{\lambda-1}$$

### 6.3 边界条件分析

边界值满足：

- $(0) = 0$
- $(2) = 0$

对任意  $D \in (0, 2)$ ，有  $(D) > 0$ 。因此最大值必出现在区间内部  $(0, 2)$ 。

端点导数：

- $(0) = 1 > 0$
- $(2) < 0$

由  $(0) > 0$ 、 $(2) < 0$  及  $(D)$  连续性，至少存在一点  $D^* \in (0, 2)$  使得  $(D^*) = 0$ 。

## 6.4 一阶条件与闭式解

令  $(D^*) = 0$ ，得：

$$\left(1 - D^*2\right) = \lambda 2D^*(D^* + )$$

整理得二次方程：

$$\lambda D^{*2} + (\lambda + 1)D^* - 2 = 0$$

正根为：

$$D^* = \frac{\sqrt{2(\lambda + 1)^2 + \lambda} - (\lambda + 1)}{2\lambda}$$

## 6.5 严格凹性与唯一性

二阶导数  $(D) < 0$  在  $\lambda > 0$ 、 $D \in (0, 2)$  上恒成立。因此  $(D)$  严格凹， $D^*$  为唯一全局最优点。

## 6.6 $D^*$ 与 $\lambda$ 的关系

临界值：

$$\lambda^* = -1$$

$\lambda$ 范围	策略类型	$D^*$ 位置
$\lambda < \lambda^*$	自利	$D^* < 1$
$\lambda = \lambda^*$	公平	$D^* = 1$
$\lambda > \lambda^*$	慷慨	$D^* > 1$

## 6.7 核心结论

**定理1（最优诉求）：**在  $(0) = 1$ 、 $(2) = 0$  及  $(D) = (1 - D2)^\lambda$  下，长期收益函数  $(D)$  在  $[0, 2]$  严格凹，存在唯一最优诉求  $D^* \in (0, 2)$ ，且：

- $\lambda < -1$   $D^* < 1$ （自利）
- $\lambda = -1$   $D^* = 1$ （公平）
- $\lambda > -1$   $D^* > 1$ （慷慨）

## 7 与传统博弈论的区别

维度	传统博弈论	收益导向博弈
核心问题	给定行为假设，均衡点在何处	何种行为最大化长期收益

维度	传统博弈论	收益导向博弈
行为假设	事前预设合作或非合作	无预设，行为内生适应
反馈机制	无	有，根据对方反应调整
退出机制	无	内含于
环境假设	确定性	随机性（含随机冲击 $\xi$ ）
收益结构	单次互动收益	单次收益×持续概率
结果类型	均衡点	动态涌现最优

传统博弈论关注“停在何处”，收益导向博弈关注“如何行动”。

## 8 意义与拓展

收益导向博弈将个体从事前行为限制中解放，使博弈论建立在适应性长期互动之上。框架揭示核心思想：**自利在长期合适互动结构中可自然涌现利他行为**。这不依赖道德，而是理性适应环境的逻辑结果。

本框架可自然拓展至交易策略、人际关系、组织管理等领域。所有重复互动均可在该框架下重新审视，为自利涌现合作提供统一视角。

## 9 结论

### 9.1 最优诉求决定机制

纳什证明：在他人不动假设下存在不动点。本文提出：在他人会行动、个体有选择、环境受随机冲击时，关键问题不是均衡是否存在，而是如何调整以最大化长期收益。

答案位于中间最优诉求  $D^*$ ，其位置由对手选择性  $\lambda$  与对方诉求 的相对关系决定。

基本原理：面对高选择性对手，慷慨是长期最优策略；面对低选择性对手，自利更有利。**利他**可以是自利的理性结果。

### 9.2 从个体最优到博弈均衡

多主体按学习规则调整策略时，在严格凹性与对称条件下，系统收敛至对称纳什均衡。本文核心贡献不在于求解均衡本身，而在于揭示均衡如何通过长期互动动态涌现。

### 9.3 哲学基础：自利致利他



本文最终哲学洞察：**自利追求在合适互动结构中可自然表现为利他行为**。这并非悖论，而是策略环境理性适应的逻辑必然。当个体面对高选择性对手时，慷慨成为长期收益最大化最优策略。表观利他并非背离理性，而是理性的高级形式。

---

## 参考文献

- [1] Nash, J. F. (1950). Equilibrium points in n-person games.
- [2] Nash, J. (1951). Non-cooperative games.
- [3] von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*.
- [4] Rubinstein, A. (1982). Perfect equilibrium in a bargaining model.
- [5] Topkis, D. M. (1998). *Supermodularity and Complementarity*.