Action Being Character: A Promising Perspective on the Solution Concept of Game Theory

Kuiying Deng[1]¤, Tianguang Chu[1,2]*

1 State Key Laboratory for Turbulence and Complex Systems, College of Engineering, Peking University, Beijing, China, 2 Key Laboratory of Machine Perception, Ministry of Education, Peking University, Beijing, China

Abstract

The inconsistency of predictions from solution concepts of conventional game theory with experimental observations is an enduring question. These solution concepts are based on the canonical rationality assumption that people are exclusively self-regarding utility maximizers. In this article, we think this assumption is problematic and, instead, assume that rational economic agents act as if they were maximizing their implicit utilities, which turns out to be a natural extension of the canonical rationality assumption. Implicit utility is defined by a player's character to reflect his personal weighting between cooperative, individualistic, and competitive social value orientations. The player who actually faces an implicit game chooses his strategy based on the common belief about the character distribution for a general player and the self-estimation of his own character, and he is not concerned about which strategies other players will choose and will never feel regret about his decision. It is shown by solving five paradigmatic games, the Dictator game, the Ultimatum game, the Prisoner's Dilemma game, the Public Goods game, and the Battle of the Sexes game, that the framework of implicit game and its corresponding solution concept, implicit equilibrium, based on this alternative assumption have potential for better explaining people's actual behaviors in social decision making situations.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: chutg@pku.edu.cn

¤ Current address: Department of Mathematics, Uppsala University, Uppsala, Sweden

Introduction

Solution concepts of conventional game theory are mainly based on the canonical rationality assumption that players are rational economic agents who act as if they were maximizing their self-regarding utilities. Nevertheless, increasing experimental data support that decision makers in human society often do not act in accordance with predictions from these kinds of solutions [1–4]. This suggests that new insights and mechanisms are required to reconcile this undeniable inconsistency between theory and experiments. Indeed, similar cases often arise as well in many other areas of study such as physics. Galileo, founder of modern mechanics, proposed that all free falling bodies would fall through a vacuum with the same uniform acceleration, and also realized that, to explain why, as legend has it, the larger one of the two iron balls he dropped from the Tower of Pisa hit the ground slightly ahead of the smaller one, the resistance from air through which the balls were falling must be taken into account [5]. Similarly, the canonical rationality assumption provides a unified launching pad for a large body of theories in economics and related fields that work well to some extent and thus should not be given up lightly, but this does not imply that this assumption can be readily applied to explain people's actual behaviors without considering the effect of real game situations they are involved in, just as that of the air resistance on free falling balls.

So far, many efforts have been devoted to the modification or extension of the canonical rationality assumption in game theory with the aid of experimental observations. For instance, bounded rationality and other-regarding preference of players are considered in behavioral game theory [4,6]. It is safe to say that people are not exclusively self-regarding [6,7]. Costly punishment across human societies even in one-shot game situations is a fine example [8,9]. People also show preferences for maximizing collective benefit or relative advantage over others' benefits in experiments [10,11]. However, almost all experiments are designed to be simple and clear to make sure that all subjects have understood the rules completely [3], so there is little sense in trying to explain these experimental data by bounded rationality. Likewise, evolutionary game theory cannot explain why there are subjects in experiments who prefer to choose the strategy that is not evolutionarily stable when they clearly know this information, such as cooperating in one-shot Prisoner's Dilemma games [12]. The solution concept based on psychological game theory invokes psychological analysis about the motivations of other players [2]. In this article we will propose a promising solution concept which can not only avoid these vulnerabilities, but also provide better explanations of people's diverse behaviors unpredicted by conventional game theory based on the canonical rationality assumption.

Although the experimental evidence that rational self-regarding utility maximizers exist in chimpanzee's societies might offer some support for the canonical rationality assumption [13], many experimental results show that people are heterogeneous and

behave differently when choosing rational strategies. The behaviors of decision makers in human society could be influenced by diverse factors such as sex [14], age [15], emotions [16], past experience [17], educational background [18], social distance [19], cultural difference [3,20], and even experimental contexts [1]. For example, economists are more likely than others to free ride in public goods experiments [18], and there are the same players who converged to contribute nothing in a ten-period public goods experiment without punishment but contribute everything in a ten-period public goods experiment with punishment [21]. "Action is character", wrote F. Scott Fitzgerald, author of The Great Gatsby, in one of his notebooks. What people do reflects what they are, and what people are implies what they do. We propose that rationality should be personal and related to the player's character. No matter how anomalous an action appears to be, it still should be thought as rational to those who take it after deliberation.

In this article we try to formally incorporate into the solution concept of game theory the player's character, a parameter encapsulating all diverse factors that may influence his decision making. The characters of players reflect their personal weighting between cooperative, individualistic, and competitive social value orientations. We assume that all players have a common belief about the character distribution for a general player, and rational players act as if they were maximizing their character-related utilities (that is, implicit utility we shall define later on), which turns out to be a natural extension of the canonical rationality assumption. In this procedure, a game is transformed into a Bayesian game, the solution of which, Bayesian Nash equilibrium, is referred to as implicit equilibrium of the original game. An implicit equilibrium provides players with a concrete decision rule which varies with their characters, and no player would feel regret about his decision even though the outcome turns out not to be a Nash equilibrium. The idea behind our framework turns out to be similar to the ERC model, a theory of equity, reciprocity, and competition [22]. But the ERC model uses equity as one of the primitive hypotheses and fails to incorporate cooperative factors, although it does consider players' heterogeneity and the tension between individualistic and competitive orientations [22]. In our framework, equity emerges in the Ultimatum game as a natural consequence of rational players acting as if they were maximizing their character-related utilities.

Solution concepts based on the canonical rational assumption also assume that it is common knowledge that each player is rational, each player knows that all players are rational, each player knows that each player knows that all players are rational, and so on *ad infinitum*. It is probably safe to say that this is not realistic in most cases. The concept of bounded rationality in behavioral game theory is proposed to avoid this vulnerability [6]. Instead, in our new framework of character and implicit equilibrium, we assume that the player facing a game situation chooses his strategy based on the common belief about the character distribution for a general player and the self-estimation of his own character, irrespective of what the characters of other players involved are, which strategies he think they will choose, and even whether they are rational or not. Rationality in our framework is only referred to as players maximizing their character-related utilities, and not as players with perfect memory and extraordinary computational capacity like supermen, and thus there is no need to invoke such concepts as bounded rationality. This contrast is one of the most important features of the concept of implicit equilibrium distinct from almost all other solution concepts proposed until now in game theory.

The concepts of character and implicit equilibrium we propose here based on this new rationality assumption can better explain people's actual behaviors labeled as irrational by conventional game theory in the experiments of some paradigmatic games. For example, why are at least one third of the dictators willing to offer the recipients more than zero in the Dictator game [23]? Why is it a modal pattern that two players end up with around fifty-fifty allocation of the pie in the Ultimatum game [3,16]? Why are there subjects who cooperate in the Prisoner's Dilemma game [1] and contribute their endowments in the Public Goods game [3]? And why do people feel well satisfied with their choices even though they do not choose their own preferred programs simultaneously in the Battle of the Sexes game? All these puzzling but attractive questions are going to be given more reasonable explanations in our new framework based on the character-related rationality assumption that is presented subsequently.

Analysis

We begin by defining an *n*-player noncooperative game in strategic form and, in passing, introducing the notations that will be used hereafter [24]. Let $I = \{1, 2, \ldots, n\}$ be the set of players, where *n* is the total number of players. For each player $i \in I$, let s_i be a strategy of his and S_i be the finite set of all his strategies. A vector $s = (s_1, s_2, \ldots, s_n)$ is called a strategy profile. The set of strategy profiles is thus the Cartesian product $S = \times_i S_i$. Let $u_i : S \to R$ be the von Neumann-Morgenstern utility function associated to player *i* and $u = (u_1, u_2, \ldots, u_n)$ be the combined utility function of the game. Then a game in strategic form may be summarized as a triplet

$$\Gamma = (I, S, u). \tag{1}$$

We decompose a player's utility into two parts: the outward-looking utility and the inward-looking utility. The former reflects the benefit of the group as a whole gained from playing the game and is the same for all players concerned, whereas the latter reflects the relative advantage of the focal player gained over the utilities of the other players involved in the game. A player's character is defined to reflect his personal weighting between the outward-looking utility and the inward-looking utility, or equivalently, his personal weighting between intergroup and intragroup competition.

Formally, for any strategy profile $s \in S$, the *outward-looking utility* of player *i* is defined as

$$u_i^{out}(s) = \frac{1}{n} \sum_{j=1}^{n} u_j(s), \tag{2}$$

and the *inward-looking utility*

$$u_i^{in}(s) = u_i(s) - u_i^{out}(s). \tag{3}$$

Based on these two new concepts, we further define the *implicit utility* of player *i* as

$$u_i^*(s) = (1 + \kappa_i) u_i^{out}(s) + (1 - \kappa_i) u_i^{in}(s), \tag{4}$$

where

$$\kappa_i \in [-1, 1] \tag{5}$$

represents player *i*'s *character* which reflects his personal weighting

between the outward-looking utility and the inward-looking utility. When $\kappa_i = 0$, $u_i^*(s)$ degenerates to $u_i(s)$. For the two-player case, $u_i^*(s) = u_i(s) + \kappa_i u_j(s)$ where $j \neq i$, and for the zero-sum game where $\sum_{j=1}^{n} u_j(s) = 0$, $u_i^*(s) = (1 - \kappa_i) u_i(s)$. So implicit utility is defined by a kind of utility transformation in principle.

A player's character may take any value in the interval $[-1,1]$ by definition. One with character $\kappa_i = 1$ cares only about the outward-looking utility and acts by team-directed reasoning to achieve a group goal [25]. In this case the allocation of utilities among players is irrelevant to him. One with character $\kappa_i = 0$ cares only about his self-regarding utility as in the canonical rationality assumption. This makes our new framework a natural extension of solution concepts in conventional game theory. One with character $\kappa_i = -1$ cares only about the inward-looking utility, that is, his advantage over other players. These three special cases reflect people's cooperative, individualistic, and competitive social value orientations, respectively, which form the theoretical basis for numerous studies [10].

Note that the inward-looking or outward-looking utility is distinct from the exclusively self-regarding or other-regarding utility. This is an issue about what factors should be considered as the primitive assumptions about people's instincts related to decision making. We think that the exclusively other-regarding preference as a primitive assumption might contradict the principle of individuals being the unit of natural selection. The definition of implicit utility in (4) is equivalent to

$$
\begin{aligned}
u_i^*(s) &= (1 - \kappa_i) u_i(s) + \frac{2\kappa_i}{n} \sum_{j=1}^{n} u_j(s) \\
&= \left[1 - \kappa_i \left(1 - \frac{2}{n} \right) \right] u_i(s) + \frac{2\kappa_i}{n} \sum_{j \neq i} u_j(s).
\end{aligned}
\tag{6}
$$

This implies that any player i, regardless of his character κ_i, always gives a positive weighting to his exclusively self-regarding utility u_i, and thus he will never show an exclusively other-regarding preference by definition. Actually, the weighting player i gives u_i is always more than that he gives u_j, where $j \neq i$, by $1 - \kappa_i$.

The character of a player, which reflects his personality when facing a game situation, may vary with the contexts in which he needs to make a decision. For example, experimental data show that the cooperation rate usually decreases in repeated public goods games without punishment relative to in those games with punishment [21], which reveals that players might update their characters by learning to adapt to particular situations. However, considering the fact that players from some specific social context usually follow the same code of conduct, whether religious, ethical, or legal, we assume that,

when facing one-shot game situations without communication, all players have a relatively stable common belief β about the character distribution for a general player over the interval $[-1,1]$.

In other words, for any focal player, β is the best of his knowledge about other players in a one-shot game situation. As far as the real world is concerned, we think that it might be inappropriate to always assume that people who face a one-shot game situation have no knowledge of the characters of other players involved.

By incorporating players' characters and their common belief about the character distribution for a general player, any game can be transformed into a Bayesian game. In particular, let $\kappa = (\kappa_1, \kappa_2, \ldots, \kappa_n)$ and $u^* = (u_1^*, u_2^*, \ldots, u_n^*)$. Then a Bayesian game

$$
\Gamma^* = (I, S, \kappa, \beta, u^*)
\tag{7}
$$

can be defined. We call Γ^* the *implicit game* corresponding to Γ, and its Bayesian Nash equilibrium,

$$
\gamma_i : [-1,1] \to S_i \quad \text{for } i = 1, 2, \ldots, n,
\tag{8}
$$

the *implicit equilibrium* of Γ, which provides the player with a concrete decision rule based on the self-estimation of his own character, irrespective of other players' characters. No matter whether the outcome turns out to be a Nash equilibrium of the game Γ or not, no player would feel regret about his decision and thus no player would want to change it.

In contrast to the canonical rationality assumption, the idea underlying this transformation is referred to as the *character-related rationality assumption*; that is,

all players are rational economic agents who act as if they were maximizing their implicit utilities.

For those players with character $\kappa_i = 0$, this is exactly the canonical rationality assumption, except that we do not need the assumption that rationality is common knowledge. We expect that the concepts of implicit game and implicit equilibrium based on this alternative assumption can better model how decisions are made in actual social contexts. Subsequently we apply these new concepts to some paradigmatic game situations to show their capability and effectiveness.

The Dictator Game

The Dictator game is about sharing a pie between two anonymous players, in which the dictator who has complete control over the process of allocation chooses to offer a portion of the pie, x, to the recipient who has no choice but to accept it, and thus leaves $1 - x$ to himself (Table 1). This one-shot game is designed to measure the altruism behavior of the dictator, which is not directly associated with other plausible factors such as kinship, reciprocity, reputation, or the immediate threat of punishment [8].

According to the canonical rationality assumption, the rational dictator who acts as if maximizing his self-regarding utility would take advantage of his dominant position and thus choose to offer the recipient nothing. However, a consistent observation in experiments is that at least one third of the dictators are willing to offer the recipients more than zero across a wide variety of procedures and conditions, although it is true that most of the dictators offer nothing [23].

In contrast, the character-related rationality assumption predicts that not all the dictators who act as if maximizing their implicit utilities would offer nothing, as shown in the following proposition.

Proposition 1. *In the Dictator game, (i) a rational dictator with character $\kappa_1 = 1$ would like to offer the recipient any portion of the pie, even as*

Table 1. The Dictator Game.

Dictator	Recipient Accept
Offer x	$1 - x, x$

big as the whole pie, and (ii) a rational dictator with character $\kappa_1 < 1$ would offer the recipient nothing.

Proof. Suppose that the dictator, player 1, offers a portion of the pie, x, to the recipient, player 2, and leaves $1-x$ to himself. The implicit utility of player 1 is

$$u_1^*(x) = (1+\kappa_1)(0.5) + (1-\kappa_1)(0.5-x)$$
$$= 1-(1-\kappa_1)x. \qquad (9)$$

If $\kappa_1 = 1$ then $u_1^*(x) \equiv 1$. That is, the implicit utility of the dictator with character $\kappa_1 = 1$ is constantly equal to 1 whatever he gets from the allocation of the pie. So a rational dictator with character $\kappa_1 = 1$ would like to offer any portion of the pie x to the recipient, even as big as the whole pie.

If $\kappa_1 \neq 1$ then $\arg\max_{x \in [0,1]} u_1(x) = 0$ and $u_1(0) = 1$. That is, a rational dictator with character $\kappa_1 < 1$ who try to maximize his implicit utility would offer the recipient nothing so as to get the maximum implicit utility 1.

As a result of the recipient's inability to influence the allocation procedure, the implicit equilibrium presented in this proposition does not even need the dictator and the recipient to have a common belief β about the character distribution for a general player.

It is worth noting that we can get $x \leq \frac{1}{1-\kappa_1}$ by solving $u_1^*(x) \geq 0$ in the proof above, which means that the dictator with character $\kappa_1 < 1$ would still be better off if he offers the recipient x not more than $\min\{1, \frac{1}{1-\kappa_1}\}$ than not participating in the game, even though he does not manage to maximize his implicit utility due to some unknown reasons. In particular, those dictators with non-negative characters who fail to maximize their implicit utilities could be better off even though they offer the whole pie. This point may be helpful to the understanding and explanation of why there are the dictators who are willing to offer more than zero in experiments, except that those dictators with character $\kappa_1 = 1$ who care only about the collective benefit would anyway.

The Ultimatum Game

The Ultimatum game is similar to the Dictator game. One only difference is that, in the Ultimatum game, after the proposer offers a portion of the pie, x, to the responder and leaves $1-x$ to himself, the responder could choose either to accept it with the allocation of the pie settled down or to reject it with two players getting nothing (Table 2) [8]. The responder in this one-shot anonymous game is endowed with an opportunity to punish the proposer's greed, but, if he did, what he does would cost him all he could own.

According to the canonical rationality assumption, this game has a subgame perfect Nash equilibrium where the proposer offers the smallest nonzero amount and the responder accepts it, because a rational proposer knows that the responder who maximizing his self-regarding utility will always accept any positive offer [8].

Table 2. The Ultimatum Game.

Proposer	Responder	
	Accept	Reject
Offer x	$1-x, x$	0, 0

However, experimental evidence shows that (i) the offer x distributes over the whole interval $[0,1]$ [3], (ii) offers above 0.5 are rare in student subjects [26], (iii) a modal pattern is that two players end up with around fifty-fifty allocation of the pie [3,16], and (iv) there are the responders who accept the smallest nonzero offer and the proposers who offer the whole pie [3].

In contrast, the character-related rationality assumption that players act as if they were maximizing their implicit utilities provides a more reasonable explanation for all these experimental observations.

Proposition 2. *In the Ultimatum game, (i) a rational proposer with character $\kappa_1 = 1$ would like to offer any portion of the pie not less than $\max\{0, \frac{\kappa_2}{\kappa_2 - 1}\}$, even as big as the whole pie, where κ_2 is the character of the responder, and the responder would accept it, and (ii) a rational proposer with character $\kappa_1 < 1$ would offer $\max\{0, \frac{\kappa_2}{\kappa_2 - 1}\}$ which is never more than half the pie, and the responder would accept it.*

Proof. Suppose that the proposer, player 1, offers a portion of the pie, x, to the responder, player 2, and leaves $1-x$ to himself. The implicit utility of player 1 is

$$u_1^*(x) = (1+\kappa_1)(0.5) + (1-\kappa_1)(0.5-x)$$
$$= 1-(1-\kappa_1)x, \qquad (10)$$

and the implicit utility of player 2 is

$$u_2^*(x) = (1+\kappa_2)(0.5) + (1-\kappa_2)(x-0.5)$$
$$= \kappa_2 + (1-\kappa_2)x. \qquad (11)$$

If an offer is accepted by both players, $u_1^*(x) \geq 0$ and $u_2^*(x) \geq 0$ are needed to be satisfied simultaneously. Solving these two inequalities, we get

$$\frac{\kappa_2}{\kappa_2 - 1} \leq x \leq \frac{1}{1-\kappa_1}. \qquad (12)$$

So an offer x that exits between $\max\{0, \frac{\kappa_2}{\kappa_2 - 1}\}$ and $\min\{1, \frac{1}{1-\kappa_1}\}$ would be accepted by both players.

Because the implicit utility $u_1^*(x)$ of a proposer with character $\kappa_1 = 1$ is constantly equal to 1 whatever he gets from the allocation of the pie, his rational strategy is to offer any x that will be accepted by the responder, that is, not less than $\max\{0, \frac{\kappa_2}{\kappa_2 - 1}\}$, even as big as the whole pie.

The implicit utility $u_1^*(x)$ of a proposer with character $\kappa_1 < 1$ increases with x decreasing, so he would like to offer the lower bound of the offer that the responder will accept so as to maximize his implicit utility, that is, $\max\{0, \frac{\kappa_2}{\kappa_2 - 1}\}$. Since $\max\{0, \frac{\kappa_2}{\kappa_2 - 1}\}$ increases with κ_2 decreasing and reaches the maximum 0.5 when $\kappa_2 = -1$, a rational proposer with character $\kappa_1 < 1$ would never offer the responder more than half the pie.

From the proof above we can also infer that the responder with non-negative character would accept any offer x by the proposer, since the lower bound $\max\{0, \frac{\kappa_2}{\kappa_2 - 1}\} = 0$ when $\kappa_2 \geq 0$. Furthermore, since the upper bound $\min\{1, \frac{1}{1-\kappa_1}\} = 1$ when $\kappa_1 \geq 0$, the proposer with non-negative character might offer the responder any x, even as big as the whole pie. In these two cases players would be better off than not participating in the game, even if they do not manage to maximize his implicit utility due to some

unknown reasons. These two points might be helpful to explain why, in experiments, the offer x distributes over the whole interval $[0,1]$ and there are the responders who accept the smallest nonzero offer and the proposers who offer the whole pie.

As in the Dictator game, the implicit equilibrium presented in this proposition does not also require the proposer and the responder to have a common belief β about the character distribution for a general player, because of the asymmetrical positions of the two players in the allocation procedure. However, close inspection will reveal that this implicit equilibrium is inconsistent with our aforementioned presentation that, when a player who acts as if maximizing his implicit utility chooses a strategy, the characters of other players are irrelevant. The strategy we give here requires the proposer to estimate the character of the responder.

Let us dig deeper into the procedure of the proof of this proposition. The upper bound of the offer that both players will accept, $\min\{1, \frac{1}{1-\kappa_1}\}$, decreases with the character of the proposer κ_1 decreasing and reaches the minimum 0.5 when $\kappa_1 = -1$, while the lower bound, $\max\{0, \frac{\kappa_2}{\kappa_2-1}\}$, increases with the character of the responder κ_2 decreasing and reaches the maximum 0.5 when $\kappa_2 = -1$. If we think of student proposers as players who care only about their inward-looking utility and are good at estimating responders' characters, then this analysis can explain why offers above 0.5 are rare in student subjects in experiments. This analysis also gives us a far-reaching conclusion as follows.

Proposition 3. *In the Ultimatum game, half the pie is the unique offer that would be always accepted by any two players.*

Now the characters of both the proposer and the responder fade out completely. Actually, in the case where $x = 0.5$, solving $u_1^*(x) \geq 0$ and $u_2^*(x) \geq 0$ gives us $\kappa_1 \geq -1$ and $\kappa_2 \geq -1$, which can be satisfied by any two players by definition. This proposition finally provides us with a perfect explanation of why around fifty-fifty allocation of the pie is observed as a modal pattern in experiments.

Fairness, or inequality aversion, is recognized as one of key elements in human strategic interactions [16]. Here we show that the proposer giving half the pie and the responder accepting it is an implicit equilibrium, irrespective of the characters of the proposer and the responder. Hence, fairness is achieved as a natural consequence of two players acting as if they were maximizing their implicit utilities, whereas in almost all other existing theoretical models it is incorporated explicitly as one of primitive assumptions about people's instincts related to decision making [10,22,26,27].

The Prisoner's Dilemma Game

The best known model in game theory might be the Prisoner's Dilemma game, which is also played between two anonymous players (Table 3). Each player in this game can either cooperate or defect. One cooperating would induce cost c to himself and bring benefit b to the other. One defecting costs nothing and benefits neither. Note that $b > c$ is usually assumed so as to make sure that two cooperating players can benefit more than two defecting players. This simple model and its finitely and infinitely repeated versions are extensively used to explore the mechanism of the evolution of cooperation through natural selection [28].

According to the canonical rationality assumption, one in this scenario would always defect since defecting is a dominant strategy. That is, one can always be better off by defecting than by cooperating, whether the other player defects or cooperates. By backward induction, we know, even in the finitely repeated

Table 3. The Prisoner's Dilemma Game.

Focal player	Another player	
	Cooperate (C)	**Defect (D)**
Cooperate (C)	$b-c$	$-c$
Defect (D)	b	0

version, defecting in all steps is the best choice. However, each could achieve a higher utility $b-c$ by both cooperating than 0 by both defecting. This is why the Prisoner's Dilemma game is often used as a synonym for "social dilemma".

The predictions given by the canonical rationality assumption contradicts experimental data. There are subjects who cooperate even in the last round of finitely repeated situations, and the experimental studies from 1958 to 1992 show that the cooperation rate of players ranges from 5% to 96.9%, with a mean of 47.4% [1]. In contrast, by the character-related rationality assumption, one with character $\kappa_i \geq \frac{c}{b}$ would cooperate in the Prisoner's Dilemma game so as to maximize his implicit utility.

Proposition 4. *In the Prisoner's Dilemma game, one would cooperate if his character $\kappa_i \geq \frac{c}{b}$, and otherwise defect, given that the two players have any common belief about the character distribution for a general player.*

Proof. For player i with character κ_i in the Prisoner's Dilemma game, when he cooperates, his implicit utility is $(1+\kappa_i)(b-c)$ if the other player cooperates and otherwise $b\kappa_i - c$ if the other player defects. When he defects, his implicit utility is $b - c\kappa_i$ if the other player cooperates and otherwise 0 if the other player defects.

We assume that the two players have a common belief β about the character distribution for a general player and the character of a player who cooperates is not less than a critical value κ_c. Then, the expected implicit utility of player i cooperating is

$$u_i^*(C) = \begin{aligned} &(1+\kappa_i)(b-c)\Pr\{\kappa_j \geq \kappa_c\} \\ &+ (b\kappa_i - c)(1 - \Pr\{\kappa_j \geq \kappa_c\}), \end{aligned} \tag{13}$$

where $\Pr\{\kappa_j \geq \kappa_c\}$ is the probability of the other player j cooperating, and the expected implicit utility of player i defecting is

$$u_i^*(D) = (b - c\kappa_i)\Pr\{\kappa_j \geq \kappa_c\}. \tag{14}$$

Solving $u_i^*(C) \geq u_i^*(D)$, we get $\kappa_i \geq \frac{c}{b}$. So $\kappa_c = \frac{c}{b}$. In other words, it is an implicit equilibrium that a player cooperates when his character $\kappa_i \geq \frac{c}{b}$, and otherwise defects.

This implicit equilibrium is symmetric in the sense that it works for either player. No player would feel regret even if he chooses to cooperate and the other player chooses to defect, although this outcome is not a Nash equilibrium. From this implicit equilibrium we can also infer that the players would tend to defect when c is near b, because the portion of people with character $\kappa_i \geq \frac{c}{b}$ becomes far less whatever the character distribution for a general player is.

Cooperation is considered as a third fundamental principle of evolution beside mutation and natural selection [29]. The repeated Prisoner's Dilemma game is used to describe cooperation between unrelated individuals. If the probability of another encounter between the same individuals exceeds the cost-to-

benefit ratio $\frac{c}{b}$, cooperation can evolve [29]. Now this proposition identifies $\frac{c}{b}$ as an important parameter in the one-shot Prisoner's Dilemma game, either. If the player self-estimates that his character exceeds $\frac{c}{b}$ then he would cooperate. These two conclusions complement each other perfectly.

The Public Goods Game

In the Public Goods game, each of n players is endowed with one unit of utility and secretly decides whether to contribute it into the public pot or to free ride. The umpire will multiply the sum of contributions in the pot by a number r greater than 1 and smaller than n, and then split the total utility equally among all players, irrespective of whether they contribute or not (Table 4) [30]. This game can be considered as an extension of the Prisoner's Dilemma game to the n-player situation [30]. In particular, when there are only two players, the Public Goods game is equivalent to the Prisoner's Dilemma game with $b = \frac{r}{2}$ and $c = 1 - \frac{r}{2}$.

According to the canonical rationality assumption, a rational player in this scenario would free ride since free riding is a dominant strategy. One will incur a loss of $1 - \frac{r}{n}$ units of utility by contributing whatever others do. However, each of n players could benefit most and get utility r from all players contributing. This makes it a social dilemma essentially the same as the Prisoner's Dilemma game.

In the Public Goods game experiments where players are permitted to contribute any fraction of their endowments, the mean contributions usually end up with between 40% and 60% [3]. Although the definition of the Public Goods game in these experiments is not exactly the same as that given above, the data surely reflects that those subjects are not exclusively self-regarding utility maximizer. By the character-related rationality assumption that players act as if they were maximizing their implicit utilities, we may give a more reasonable solution to the Public Goods game defined here.

Proposition 5. *In the Public Goods game, one would contribute his endowment if his character $\kappa_i \geq \frac{n-r}{n+r-2}$, and otherwise he would free ride, given that all players have any common belief about the character distribution for a general player.*

Proof. We use C and D to denote two strategies of the player, contributing his endowment and free riding, respectively. For any given player $i_0 \in I$, when there are m players using strategy C, the outward-looking utility of player i_0 using strategy D is

$$
\begin{aligned}
u_{i_0}^{out}(D,m) &= \frac{1}{n}\left[(n-m)\frac{mr}{n} + m\left(\frac{mr}{n}-1\right)\right] \\
&= \frac{m}{n}(r-1),
\end{aligned}
\tag{15}
$$

and the inward-looking utility is

$$
u_{i_0}^{in}(D,m) = \frac{mr}{n} - u_{i_0}^{out}(D,m) = \frac{m}{n}.
\tag{16}
$$

Table 4. The Public Goods Game.

Focal player	Other $n-1$ players m contribute (C)
Contribute (C)	$(m+1)r/n - 1$
Free ride (D)	mr/n

doi:10.1371/journal.pone.0019014.t004

Then we can work out the implicit utility of player i_0 using strategy D when there are m players using strategy C, which is

$$
\begin{aligned}
&u_{i_0}^*(D,m) \\
=\ & (1+\kappa_{i_0})u_{i_0}^{out}(D,m) + (1-\kappa_{i_0})u_{i_0}^{in}(D,m) \\
=\ & \frac{mr}{n} + \frac{m(r-2)}{n}\kappa_{i_0}.
\end{aligned}
\tag{17}
$$

We assume that all n players have a common belief β about the character distribution for a general player and the character of a player using strategy C is not less than a critical value κ_c. Hence the expected implicit utility of player i_0 using strategy D is

$$
u_{i_0}^*(D) = \sum_{m=0}^{n-1}\binom{n-1}{m}p^m(1-p)^{n-1-m}u_{i_0}^*(D,m),
\tag{18}
$$

where $p = \Pr\{\kappa_i \geq \kappa_c\}$ is the probability of any player i using strategy C.

For any given player $i_0 \in I$, when there are m players using strategy C, the outward-looking utility of player i_0 using strategy C is

$$
\begin{aligned}
u_{i_0}^{out}(C,m) &= \frac{1}{n}\left\{(n-1-m)\frac{(m+1)r}{n}\right. \\
&\left. + (m+1)\left[\frac{(m+1)r}{n}-1\right]\right\} \\
&= \frac{m+1}{n}(r-1),
\end{aligned}
\tag{19}
$$

and the inward-looking utility is

$$
\begin{aligned}
u_{i_0}^{in}(C,m) &= \frac{(m+1)r}{n} - 1 - u_{i_0}^{out}(C,m) \\
&= \frac{m+1}{n} - 1.
\end{aligned}
\tag{20}
$$

Then we can work out the implicit utility of player i_0 using strategy C when there are m players using strategy C, which is

$$
\begin{aligned}
&u_{i_0}^*(C,m) \\
=\ & (1+\kappa_{i_0})u_{i_0}^{out}(C,m) + (1-\kappa_{i_0})u_{i_0}^{in}(C,m) \\
=\ & \left(\frac{mr}{n} + \frac{r}{n} - 1\right) + \left[\frac{m(r-2)}{n} + \frac{r-2}{n} + 1\right]\kappa_{i_0}.
\end{aligned}
\tag{21}
$$

Hence the expected implicit utility of player i_0 using strategy C is

$$
u_{i_0}^*(C) = \sum_{m=0}^{n-1}\binom{n-1}{m}p^m(1-p)^{n-1-m}u_{i_0}^*(C,m).
\tag{22}
$$

For solving $u_{i_0}^*(C) \geq u_{i_0}^*(D)$, we calculate

$$
\begin{aligned}
u_{i_0}^*(C) - u_{i_0}^*(D) &= \left[\left(\frac{r}{n}-1\right) + \kappa_{i_0}\left(\frac{r-2}{n}+1\right)\right] \\
&\quad \times \sum_{m=0}^{n-1}\binom{n-1}{m}p^m(1-p)^{n-1-m} \\
&= \left(\frac{r}{n}-1\right) + \kappa_{i_0}\left(\frac{r-2}{n}+1\right).
\end{aligned}
\tag{23}
$$

Then we get $\kappa_{i_0} \geq \dfrac{n-r}{n+r-2}$ by solving $u_{i_0}^*(C) - u_{i_0}^*(D) \geq 0$. So $\kappa_c = \dfrac{n-r}{n+r-2}$. That is, it is an implicit equilibrium that a player would contribute his endowment if his character $\kappa_i \geq \dfrac{n-r}{n+r-2}$, and otherwise he would free ride.

The implicit equilibrium in this proposition is also symmetric and works for all players. For the two-player case, the critical value $\dfrac{n-r}{n+r-2}$ is consistent with $\dfrac{c}{b}$ in the Prisoner's Dilemma game with $b = \dfrac{r}{2}$ and $c = 1 - \dfrac{r}{2}$. Note that the limit of $\dfrac{n-r}{n+r-2}$ is 1 as r approaches 1 or n is very large relative to r, so the probability of players contributing his endowment is extremely low. In other words, the player tends to free ride when the multiplication factor r is small or the number of the players involved, n, is large.

Note that the implicit equilibria of the Dictator game and the Ultimatum game are not related to the players' common belief β about the character distribution for a general player. The implicit equilibria of the Prisoner's Dilemma game and the Public Goods game do, but not to what the common belief β itself is. In general these are not the case, for instance, as shown below in the Battle of the Sexes game.

The Battle of the Sexes Game

In the Battle of the Sexes game [2], one of two anonymous players prefers to watch soccer, whereas the other prefers to watch opera. Each player can get one unit of utility by watching his preferred program and additional x units by getting the other for company (Table 5). When $x > \dfrac{1}{2}$, these two players as a whole will be better off by watching the same program, either soccer or opera, than by watching different ones. This game is a formal representation of coordination problems which are widespread in life, such as choosing which side of the road to drive and what gifts to exchange like the young couple in the story of The Gift of the Magi by O. Henry. Another well-known game that essentially has the same coordination nature as the Battle of the Sexes game is the Stag Hunt game [2].

According to the canonical rationality assumption, this game has two pure strategy Nash equilibria: both watching soccer or both watching opera. In addition, this game has a mixed strategy Nash equilibrium; that is, each player watches his preferred program with probability $\dfrac{1}{2}\left(1 + \dfrac{1}{x}\right)$ and the other player's preferred program with probability $\dfrac{1}{2}\left(1 - \dfrac{1}{x}\right)$. When this game is played once without communication, multiple Nash equilibria can not give players much instruction, and disequilibrium may result. In contrast, the character-related rationality assumption that players act as if they were maximizing their implicit utilities can provide players with a more concrete strategy of play, given that they have a common belief β about the character distribution for a general player.

Table 5. The Battle of the Sexes Game.

Focal player	Another player	
	Opera (Own)	Soccer (Other)
Soccer (Own)	1	$1+x$
Opera (Other)	x	0

Proposition 6. *In the Battle of the Sexes game, one would watch the other player's preferred program if his character* $\kappa_i \geq \dfrac{1}{2}\left(\sqrt{1 + \dfrac{4}{x}} - 1\right)$, *and otherwise his own preferred program, given that both players believe that the character of a general player is uniformly distributed over the interval* $[-1, 1]$.

Proof. For two players, i and j, in the Battle of the Sexes game, when player i watches his preferred program, his implicit utility is $(1+x) + \kappa_i x$ if player j accompanies him and otherwise $1 + \kappa_i$ if player j watches his own preferred program. When player i watches player j's preferred program, his implicit utility is $x + \kappa_i(1+x)$ if player j accompanies him and otherwise 0 if player j watches player i's preferred program.

We assume that the two players have a common belief β about the character distribution for a general player and player i would watch player j's preferred program when $\kappa_i \geq \kappa_c$. The expected implicit utility of player i watching the other player j's preferred program is

$$u_i^*(Other) = (x + \kappa_i + \kappa_i x)(1 - \Pr\{\kappa_j \geq \kappa_c\}), \qquad (24)$$

where $\Pr\{\kappa_j \geq \kappa_c\}$ is the probability of player j watching player i's preferred program. The expected implicit utility of player i watching his own preferred program is

$$u_i^*(Own) = \begin{aligned} &(1 + x + \kappa_i x)\Pr\{\kappa_j \geq \kappa_c\} \\ &+ (1 + \kappa_i)(1 - \Pr\{\kappa_j \geq \kappa_c\}). \end{aligned} \qquad (25)$$

Solving $u_i^*(Other) \geq u_i^*(Own)$ gives us

$$x(1 - 2\Pr\{\kappa_j \geq \kappa_c\})\kappa_i \geq 1 - x + 2x\Pr\{\kappa_j \geq \kappa_c\}. \qquad (26)$$

Further we assume that both players believe that the character of a general player is uniformly distributed over the interval $[-1, 1]$, which leads to $\Pr\{\kappa_j \geq \kappa_c\} = \dfrac{1 - \kappa_c}{2}$. Substituting it into the above inequality gives us

$$(\kappa_i + 1)\kappa_c x \geq 1. \qquad (27)$$

Considering that $x > 0$ and $\kappa_i \in [-1, 1]$, one necessary condition for this inequality to be justifiable is $\kappa_c > 0$. So we can get

$$\kappa_i \geq \dfrac{1}{\kappa_c x} - 1. \qquad (28)$$

Setting $\dfrac{1}{\kappa_c x} - 1 = \kappa_c$ and solving it give us $\kappa_c = \dfrac{1}{2}\left(\sqrt{1 + \dfrac{4}{x}} - 1\right)$. Thus it is an implicit equilibrium that one would watch the other player's preferred program if his character $\kappa_i \geq \dfrac{1}{2}\left(\sqrt{1 + \dfrac{4}{x}} - 1\right)$, and otherwise he would watch his own preferred program.

The procedure of proof shows that the implicit equilibrium in the Battle of the Sexes game is closely related to the players' common belief β about the character distribution for a general player. As an alternative, if we assume that both players believe that the character of a player is uniformly distributed over the interval $[0, 1]$, then $\Pr\{\kappa_j \geq \kappa_c\} = 1 - \kappa_c$. Then the same procedure as in the proof above leads to $\kappa_c = \dfrac{1}{4}\left(\sqrt{9 + \dfrac{8}{x}} - 1\right)$. Consequently, we arrive at the following conclusion.

Proposition 7. *In the Battle of the Sexes game, one would watch the other player's preferred program if his character* $\kappa_i \geq \frac{1}{4}\left(\sqrt{9+\frac{8}{x}}-1\right)$, *and otherwise his own preferred program, given that both players believe that the character of a general player is uniformly distributed over the interval* $[0,1]$.

Under both assumptions about the players' common belief β about the character distribution for a general player, the implicit equilibria are symmetric and works for either player. Further, we can expect that one will watch the other player's preferred program with probability $\frac{1}{4}\left(3-\sqrt{1+\frac{4}{x}}\right)$ in the first case, and with probability $\frac{1}{4}\left(5-\sqrt{9+\frac{8}{x}}\right)$ in the second case. Yet we have to emphasize that no player feels regret in implicit equilibrium even if they watch different programs, that is, they simultaneously choose to watch their own or the other player's preferred programs. This is totally distinct from Nash equilibrium where both players watching different programs is considered as disequilibrium in which they feel regret and want to change their strategies.

When additional utility x from the company of the other player is extremely large (or extremely small), the implicit equilibria in both cases predict that the probability of one watching the other player's preferred program would approach $\frac{1}{2}$ (or 0). These are consistent with the predictions of the mixed strategy Nash equilibrium which says this probability is $\frac{1}{2}\left(1-\frac{1}{x}\right)$. Nevertheless, when x is between $\frac{1}{2}$ and 1, the three Nash equilibria degenerate to one where each player would watch his own preferred program with certain, although these two players as a whole can still be better off by both watching soccer or both watching opera. In contrast, we can expect that one will watch the other player's preferred program with positive probability by the predictions of the implicit equilibria under both assumptions above. This point adds to the evidence that the implicit equilibrium might be more reasonable than the three Nash equilibria in the Battle of the Sexes game.

Results and Discussion

The inconsistency of predictions from solution concepts of conventional game theory with experimental observations is an enduring question. By clearly taking the player's heterogeneity into consideration, we have shown in five paradigmatic games, the Dictator game, the Ultimatum game, the Prisoner's Dilemma game, the Public Goods game, and the Battle of the Sexes game, that the concepts of implicit game and implicit equilibrium have potential for better explaining people's actual behaviors in social decision making situations. This theoretical framework is based on the so-called character-related rationality assumption that rational economic agents act as if they were trying to maximize their implicit utilities. The cooperative, individualistic, and competitive social value orientations are incorporated as three special cases where the player's character equals 1, 0, and -1, respectively [10]. The player chooses his strategy based on the common belief about the character distribution for a general player and the self-estimation of his own character, and he is not concerned about which strategies other players will choose and will never feel regret about his decision.

The concept of character reflects players' heterogeneity, although it does not explicitly consider complicated psychological motivations as in psychological game theory [2]. The ERC model does consider players' heterogeneity and the tension between individualistic and competitive orientations, but fails to incorporate cooperative factors [22]. The concept of implicit utility is basically a kind of utility transformation, but it is based on the player's self-estimation of his own character, rather than on some inflexible parameters that are supposed to be imposed on all players as in other models [10]. We suppose that the character-related rationality assumption as a natural extension of the canonical rationality assumption is the consequence of evolution, and surely we do not explicitly consider its evolutionary process as in evolutionary game theory [12] or bounded rationality as in behavioral game theory [6].

However, as aforementioned, a player's character could be influenced by so many factors that it seems to be a big question how to measure it. Even if this is the case, incorporating the player's character in analyzing games would no doubt be useful to explain people's diverse behaviors in decision making as we shown here. At this point it may be suitable to mention another fact from physics. The temperature of an object is also influenced by many complicated factors, yet now we can conveniently measure it using a thermometer. Likewise, constructing a proper procedure, or even a convenient device, to measure the character of a player facing a game situation is a promising and challenging project. Another interesting issue is how the character of a player (consequently his strategy) evolves in repeated game situations with new information about the strategies of other players in previous plays. The progress in either direction would further extend our understanding about the reason people cooperate in social dilemmas and the way decisions are actually made.

Acknowledgments

Author Contributions

Conceived the research: KD. Performed the research: KD TC. Wrote the paper: KD TC.

References

1. Sally D (1995) Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. Ration Soc 7: 58–92.
2. Colman AM (2003) Cooperation, psychological game theory, and limitations of rationality in social interaction. Behav Brain Sci 26: 139–198.
3. Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, et al. (2005) "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. Behav Brain Sci 28: 795–855.
4. Gintis H (2007) A framework for the unification of the behavioral sciences. Behav Brain Sci 30: 1–61.
5. Adler CG, Coulter BL (1978) Galileo and the Tower of Pisa experiment. Am J Phys 46: 199–201.
6. Camerer CF, Fehr E (2006) When does "economic man" dominate social behavior? Science 311: 47–52.
7. Fehr E, Fischbacher U (2003) The nature of human altruism. Nature 425: 785–791.
8. Henrich J, McElreath R, Barr A, Ensminger J, Barrett C, et al. (2006) Costly punishment across human societies. Science 312: 1767–1770.
9. Herrmann B, Thöni C, Gächter S (2008) Antisocial punishment across societies. Science 319: 1362–1367.
10. Van Lange PAM (1999) The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. J Pers Soc Psychol 77: 337–349.
11. Rose J, Colman AM (2007) Collective preferences in strategic decisions. Psychol Rep 101: 803–815.
12. Nowak MA, Sigmund K (2004) Evolutionary dynamics of biological games. Science 303: 793–799.
13. Jensen K, Call J, Tomasello M (2007) Chimpanzees are rational maximizers in an ultimatum game. Science 318: 107–109.
14. Eckel CC, Grossman PJ (1998) Are women less selfish than men?: Evidence from dictator experiments. Econ J 108: 726–735.

15. Fehr E, Bernhard H, Rockenbach B (2008) Egalitarianism in young children. Nature 454: 1079–1083.

16. Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003) The neural basis of economic decision-making in the ultimatum game. Science 300: 1755–1757.

17. Ahn TK, Ostrom E, Schmidt D, Shupp R, Walker J (2001) Cooperation in PD games: Fear, greed, and history of play. Public Choice 106: 137–155.

18. Frank RH, Gilovich T, Regan DT (1993) Does studying economics inhibit cooperation? J Econ Pers 7: 159–171.

19. Hoffman E, McCabe K, Smith VL (1996) Social distance and other-regarding behavior in dictator games. Am Econ Rev 86: 653–660.

20. Roth AE, Prasnikar V, Okuno-Fujiwara M, Zamir S (1991) Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. Am Econ Rev 81: 1068–1095.

21. Fehr E, Gintis H (2007) Human motivation and social cooperation: Experimental and analytical foundations. Annu Rev Sociol 33: 43–64.

22. Bolton GE, Ockenfels A (2000) ERC: A theory of equity, reciprocity, and competition. Am Econ Rev 90: 166–193.

23. Murnighan JK, Oesch JM, Pillutla M (2001) Player types and self-impression management in dictatorship games: Two experiments. Game Econ Behav 37: 388–414.

24. Weibull JW (1995) Evolutionary Game Theory MIT Press, Cambridge, Mass.

25. Sudgen R (2000) Team preferences. Econ Philos 16: 175–204.

26. Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. Q J Econ 114: 817–868.

27. Rabin M (1993) Incorporating fairness into game theory and economics. Am Econ Rev 83: 1281–1302.

28. Doebeli M, Hauert C (2005) Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game. Ecol Lett 8: 748–766.

29. Nowak MA (2006) Five rules for the evolution of cooperation. Science 314: 1560–1563.

30. Hauert C, Holmes M, Doebeli M (2006) Evolutionary games and population dynamics: maintenance of cooperation in public goods games. Proc R Soc B 273: 2565–2570.