Contributions to Game Theory and Management, XV, 250–264

Concept of Social Value Orientation in Measuring Cooperative Behavior Incentive in Games*

Yuyan Xia[1] **and Yin Li**[1,2]

[1] *St. Petersburg State University,*
Faculty of Applied Mathematics and Control Processes,
7/9, Universitetskaya nab., St. Petersburg, 199034, Russia
[2] *School of Mathematics, Harbin Institute of University,*
92, West Dazhi St., Harbin, 15000, China
E-mail: 154110443@qq.com*;* liyinrus@outlook.com

Abstract The purpose of this study is to explore the effects and mechanisms of cooperation indicators such as interpersonal trust, social preferences, and social value orientations on cooperative behavior, and we develop a psychologically grounded model of trust-based cooperation. Some accounts of cooperation in product choice games have focused on developing simple indicators of game severity i.e., the extent to which the game facilitates noncooperative choices that are derived exclusively from the game's payoff structure. However, this approach does not provide a clear explanation of the psychological mechanisms why the game's payoffs affect the rate of cooperation. When considering individuals' social preferences and positive expectations (beliefs) for how to predict the emergence of trust-based cooperation as an expected utility maximizing strategy, we show a simple decision model for trust-based cooperation. In addition, we show how these predictions are associated with a game-specific cooperation index. We then describe under what conditions trust-based cooperation is rationalizable and how trust-based decisions can be understood through the interplay between of payoffs, preferences, and beliefs.

Keywords: product choice game, cooperative behavior, K-index, social preference, trust, social value orientation.

1. Introduction

Empirical evidence suggests a systematic relationship between the cooperation rate and the payoff structure of the product choice (PC) game (Glöckner and Hilbig, 2012; Rapoport and Chammah, 1965; Steele Tedeschi, 1967; Vlaev and Chater, 2006), which has led researchers to design metrics that predict the overall cooperation rate from the payoff structure of the game (see Figure 2). Arguably the best known indicator of this type is Rapoport's cooperation K-index (1967), but other indicators are also available (e.g., Axelrod, 1967; Roth and Murnighan, 1978).The Rapoport index is based on two considerations: (i) the higher the potential payoffs from cooperation (i.e., B and A), the higher the expected cooperation rate; and (ii) the lower the potential payoffs from betrayal (i.e., C and D), the lower the expected cooperation rate. the K-index is calculated by taking the difference between the decision maker's (DM's) best payoff from cooperation and DM's worst payoff from betrayal and the worst returns: $K = \frac{B-A}{C-D}$,Thus, the K-index reflects, at least to some extent, the severity of the product choice game (we use the term severity,

*This work was supported by the China Scholarship Council grant No. 202109010045.
https://doi.org/10.21638/11701/spbu31.2022.19

which is consistent with Rapoport and Chumar's (1965) definition to refer to the general temptation to defect). the higher the K-index, the less severe the dilemma, and so the higher the cooperation rate is expected to be, all other things being equal.

However, the severity of the product choice game as its payoff can only have an impact on the behavior of the decision maker if the product choice game has positive other preferences (i.e., the decision maker derives some positive utility from the payoffs of the other players). In addition, the choice of cooperation will also depend on cooperation given the specific payoff structure of the product choice game and the decision maker's particular level of attention to the payoffs of other players. That is, the potential determinants of cooperation in a product choice game, in the context of a particular payoff structure, are preferences, beliefs, and trust. To accurately predict the rate of cooperation in a product choice game, the interaction between these three factors must be considered.

This last statement provides the central insight for this paper. Using a simple model of DMs' social preferences, and their beliefs about other players' expected choices, we use this three factors to predict when DMs will choose trust and thus take cooperative actions in a one-time PC game, given the game's specific reward structure. Furthermore, we show how different indices of cooperation can be extracted from such models and how they relate to the K-index of cooperation and to each other. Although these summary indices are useful, the psychological factors responsible for trust-based cooperation are the most intriguing.

2. The Product Choice Game Model

2.1. The Product Choice Game

		2	
		h	ℓ
1	H	2, 3	0, 2
	L	3, 0	1, 1

Fig. 1. The simple Product Choice Game

Consider the game shown in Figure 1. You can think of Player 1 as a stationary firm capable of either high effort (H) or low effort (L) in the production of its output. Player 2 refers to consumers who can buy high priced product (h) or low-priced product (l). For example, we might think of player 1 as a restaurant with elegant dinners and burgers on the menu, or a surgeon who treats respiratory problems with heart surgery or folk remedies.

Player 2 prefers high-priced products if the company puts in a lot of effort, but if not, player 2 prefers low-priced products. People may prefer fine dinners or heart surgery performed by a high-effort chef or doctor to fast food or ineffective but unobtrusive treatments. The company prefers consumers to buy high-priced products and is willing to put in a high degree of effort to induce consumer choice.

However, in a game that moves simultaneously, the consumer cannot visibly choose the effort before the consumer chooses the product. Because high effort is expensive, stationary companies prefer low effort, regardless of the consumer's choice.

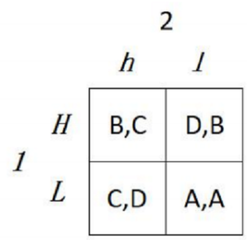


Fig. 2. The standard Product Choice Game

This work converts the product choice game in Figure 1 into a standard 2Γ—2 product choice game (Figure 2), where the payoff satisfies the strict inequality $C > B > A > D \geq 0$. See Figure 2, in PC Game, player 1 (firm) can choose between high investment (H) and low investment (L), player 2 (customer) can choose to buy high-priced products (h) or low-priced products (l). They also have two strategies to choose, then the strategy combinations of the player and the opponent are: Hh, Hl, Lh, Ll and hH, hL, lH, lL. The usual form of product choice game is instantiated when the payoff conforms to the strict inequality $C > B > A > D \geq 0$. To focus, let's anchor $C = 1$ and $D = 0$ for all games. Also, let's restrict B and A to be divisible by 0.1. This reduces the number of PC Games we will consider, but does so without loss of generality and evenly covers the space of possible PC Games. This discrete configuration produces 32 and 20 different PC Games for the company and customers, respectively. The games are shown in Table 1 and Table 2, and the corresponding K-index for each game, as well as other summary indices, will be explained in more detail later in this paper. Note that different PC Games can have the same K-index.

2.2. Social Preference

There is ample evidence that people are heterogeneous in the way they assess joint benefits (Van Lange, 1999), and that other relevant preferences can be rationalized in a utility framework (e.g., Andreoni and Miller, 2002). The most basic representation of social preference can be achieved by the decision maker's joint utility function, which appends a single parameter (α) to the returns of other actors:

$$u(\pi_s, \pi_0) = \pi_s + \alpha \cdot \pi_0 \tag{1}$$

Where, π_s is the payoff of DM (self-payoff), and π_0 is the payoff of other players. α is an index of preference for others, which is consistent with the concept of social value orientation. When α is equal to 0, narrow self-interests can be satisfied in this framework.

Table 1. Standard Product Choice Game Types (for Player 1)

PC Game	C	B	A	D	K-index	CoopArea[1]	α_{2-crit}[2]	PoA[3]
1	1	0.2	0.1	0	0.1	0.17	0.82	6.00
2	1	0.3	0.2	0	0.1	0.15	0.82	3.25
3	1	0.4	0.3	0	0.1	0.17	0.82	2.33
4	1	0.5	0.4	0	0.1	0.20	0.82	1.88
5	1	0.6	0.5	0	0.1	0.23	0.82	1.60
6	1	0.7	0.6	0	0.1	0.25	0.82	1.42
7	1	0.8	0.7	0	0.1	0.28	0.82	1.29
8	1	0.9	0.8	0	0.1	0.31	0.82	1.19
9	1	0.3	0.1	0	0.2	0.35	0.67	6.50
10	1	0.4	0.2	0	0.2	0.30	0.67	3.50
11	1	0.5	0.3	0	0.2	0.32	0.67	2.50
12	1	0.6	0.4	0	0.2	0.29	0.67	2.00
13	1	0.7	0.5	0	0.2	0.31	0.67	1.70
14	1	0.8	0.6	0	0.2	0.33	0.67	1.50
15	1	0.9	0.7	0	0.2	0.36	0.67	1.36
16	1	0.4	0.1	0	0.3	0.48	0.54	7.00
17	1	0.5	0.2	0	0.3	0.45	0.54	3.75
18	1	0.6	0.3	0	0.3	0.41	0.54	2.67
19	1	0.7	0.4	0	0.3	0.40	0.54	2.13
20	1	0.8	0.5	0	0.3	0.41	0.54	1.80
21	1	0.9	0.6	0	0.3	0.42	0.54	1.58
22	1	0.5	0.1	0	0.4	0.59	0.43	7.5
23	1	0.6	0.2	0	0.4	0.56	0.43	4.00
24	1	0.7	0.3	0	0.4	0.54	0.43	2.83
25	1	0.8	0.4	0	0.4	0.52	0.43	2.25
26	1	0.9	0.5	0	0.4	0.51	0.43	1.90
27	1	0.6	0.1	0	0.5	0.68	0.33	8.00
28	1	0.7	0.2	0	0.5	0.66	0.33	4.25
29	1	0.8	0.3	0	0.5	0.65	0.33	3.00
30	1	0.9	0.4	0	0.5	0.63	0.33	2.38
31	1	0.7	0.1	0	0.6	0.76	0.25	8.50
32	1	0.8	0.2	0	0.6	0.74	0.25	4.50
33	1	0.9	0.3	0	0.6	0.73	0.25	3.17
34	1	0.8	0.1	0	0.7	0.83	0.18	9.00
35	1	0.9	0.2	0	0.7	0.78	0.18	4.75
36	1	0.9	0.1	0	0.8	0.89	0.11	9.50

Note: For the firm, these are all possible PC games with $C = 1$, $D = 0$, B and A step 0.1, and satisfy $C > B > A > D \geq 0$.

[1] CoopArea refers to the area of the cooperative area.

[2] α_{2-crit} refers to critical α_2, which is the threshold value (also called threshold) of whether player 2 chooses to cooperate, see Eq.(9).

[3] PoA refers to "Price of Anarchy",and $PoA = \frac{B+C}{2A}$.

Table 2. Standard Product Choice Game Types (for Player 2)

PC Game	C	B	A	D	K-index	CoopArea	α_{1-crit}[4]	PoA
1	1	0.6	0.5	0	0.1	0.73	0.09	1.60
2	1	0.7	0.6	0	0.1	0.61	0.27	1.42
3	1	0.8	0.7	0	0.1	0.5	0.45	1.29
4	1	0.9	0.8	0	0.1	0.42	0.64	1.19
5	1	0.6	0.4	0	0.2	0.83	0.00	2.00
6	1	0.7	0.5	0	0.2	0.71	0.17	1.70
7	1	0.8	0.6	0	0.2	0.58	0.33	1.50
8	1	0.9	0.7	0	0.2	0.47	0.50	1.36
9	1	0.7	0.4	0	0.3	0.8	0.08	2.13
10	1	0.8	0.5	0	0.3	0.69	0.23	1.80
11	1	0.9	0.6	0	0.3	0.55	0.38	1.58
12	1	0.7	0.3	0	0.4	0.89	0.00	2.83
13	1	0.8	0.4	0	0.4	0.79	0.14	2.25
14	1	0.9	0.5	0	0.4	0.65	0.29	1.90
15	1	0.8	0.3	0	0.5	0.87	0.07	3.00
16	1	0.9	0.4	0	0.5	0.76	0.20	2.38
17	1	0.8	0.2	0	0.6	0.94	0.00	4.50
18	1	0.9	0.3	0	0.6	0.85	0.13	3.17
19	1	0.9	0.2	0	0.7	0.92	0.06	4.75
20	1	0.9	0.1	0	0.8	0.97	0.00	9.50

Note: For the customer, these are all possible PC games with $C = 1$, $D = 0$, B and A step 0.1, and satisfy $C > B > A > D \geq 0$.

[4] α_{1-crit} refers to critical α_1, which is the threshold value (also called threshold) of whether player 1 chooses to cooperate, see Eq.(8).

2.3. Beliefs: Positive expectations about the other person

In this paper, we assume that DM believes that other players will choose strategy H (or h) with probability β. If DM determines that the other players will cooperate, then β is equal to 1; conversely, if DM determines that the other players will betray, then β is equal to 0. The gradient between these two extremes is captured by different beta values in the probability space from $[0, 1]$. The standard normative model assumes that DM is confident that other players will never choose strategy C. DM may have some non-zero expectations of other players for models that have been developed before. Probably the most famous work on this route is Kreps, Milgrom, Roberts and Wilson (1982).

2.4. Trust

In this paper, we argue that DM's choice of cooperation in PC game is an expression of trust, because PC game is essentially a simple trust game. More specifically, PC game is a two-player, two-option, asymmetrical, simultaneous game of trust. (see Berg, Dickhaut, and Mc Cabe, 1995). Along these lines, the choice of collaboration shows positive intentions and positive expectations. This view is consistent with the well-known definition of trust. Take Rousseau, Sitkin, Burt, and Camerer (1998, p.395) as an example: "Trust is a state of mind that includes an intention to accept vulnerability based on positive expectations about the intentions or actions of others". We suggest adding the following to this definition: "...with the aim of improving collective outcomes". This is a useful addition because it emphasizes that trust is an intentional choice, and that in choosing to cooperate, DM has some pro-social preferences and a goal of promoting collective efficiency, which is valued by DM. In addition, the appendix to this definition explains why a DM would willingly take the strategic risk of being exploited by other players. In our view, the reason is that DM must have some other preference and a combination of positive expectations to justify this choice.

When DM chooses to cooperate in a one-shot PC game, she chooses to accept the vulnerability of being betrayed for the D payoff. The motivation for this is full confidence that the other players will choose to cooperate (β) and that the marginal improvement ($B - D$) is sufficient to warrant that this strategy "risks" over the strictly "safer" alternative (C or A). This approach rationalizes trust-based cooperative decision-making.

The question that ensues is: what combinations of prosocial preferences, positive expectations, and available payoffs would motivate DM to trust and therefore choose to cooperate in social dilemmas?

3. A Trust-based Cooperation Model

We can predict whether DM will choose to cooperate in the PC game by finding the maximum value of the expected utility of the two strategies given the DM's preferences, beliefs. Among them, α_1 and α_2 are social preferences (SVO) of players 1 and 2 respectively, β_1 and β_2 are beliefs of players 1 and 2 respectively, and the payoff structure ($CBAD$) of the PC game. In this case:

Expected utility for player 1 (firm) to choose high effort:

$$U(H) = \beta_1(B + \alpha_2 C) + (1 - \beta_1)(D + \alpha_2 B) \tag{2}$$

Expected utility for player 1 (firm) to choose low effort:

$$U(L) = \beta_1(C + \alpha_2 D) + (1 - \beta_1)(A + \alpha_2 A) \tag{3}$$

Expected utility for player 2 (customer) to choose high price product:

$$U(h) = \beta_2(\alpha_1 B + C) + (1 - \beta_2)(\alpha_1 C + D) \tag{4}$$

Expected utility for player 2 (customer) to choose low price product:

$$U(l) = \beta_2(\alpha_1 D + B) + (1 - \beta_2)(\alpha_1 A + A) \tag{5}$$

When $U(H)$ is strictly greater than $U(L)$ or $U(h)$ is strictly greater than $U(l)$, DM will choose the high input. Given this representation, we can determine the critical values of α_1, α_2, β_1 and β_2, that form the threshold between high and low engagement as a subjective expected utility maximization strategy.

4. Results

The Figure 3 shows when trust-based cooperation is required in a PC game with $C = 1$, $B = 0.8$, $A = 0.6$, $D = 0$. It represents the different positive expectations of cooperation between different SVOs and other participants on the x and y axes. The gray area indicates when a company is best to respond with low engagement, while the white area indicates when a company is best to respond with

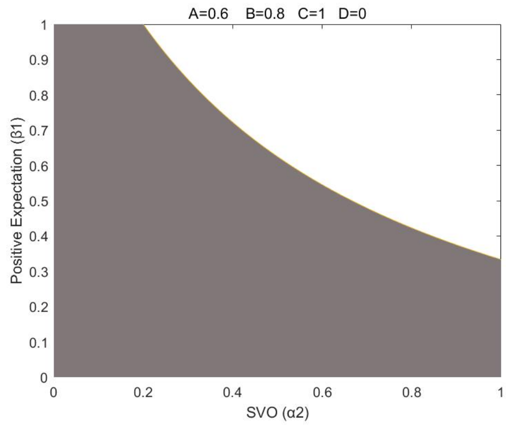


Fig. 3. Trust-Based Collaboration Figure (for Player1)

high engagement. As expected, DM with high SVO and high positive expectations would choose high engagement. The idealized DM economicus (homo economicus) is just a special case, it is located at the origin, α_2 and β_1 are 0.

The Figure 4 shows when customers need trust-based cooperation in a PC game with $C = 1$, $B = 0.8$, $A = 0.6$, $D = 0$. The dark area indicates when purchasing a low priced product is a best response, whereas the light area indicates when purchasing a high priced product is a best response. As expected, DM with high SVO and high positive expectations would choose high engagement.

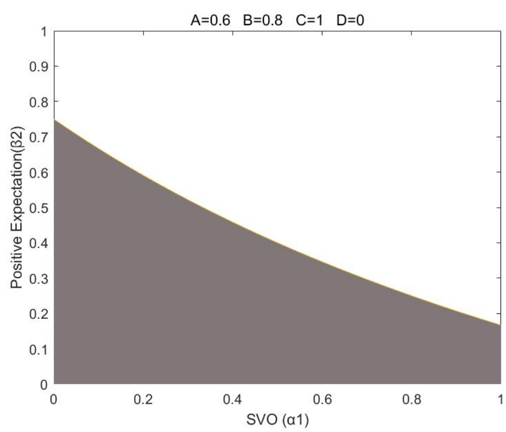


Fig. 4. Trust-Based Collaboration Figure (for Player2)

Figure 3 and Figure 4 show a special PC game with returns $C = 1$, $B = 0.8$, $A = 0.6$ and $D = 0$, respectively. The figure shows the combination of alpha and beta values for the interval [0,1] where one would expect decision makers to cooperate in a particular PC game. The light-colored coordinates on the Cartesian plane represent the case of $U(H) > U(L)$ (or $U(h) > U(l)$), and thus cooperative predictions. Conversely, the dark coordinates represent the case of $U(H) < U(L)$ (or $U(h) < U(l)$), indicating non-cooperation.

Given a completely selfish DM, we wouldn't expect cooperation, no matter what DM thinks the other players will choose. However, if DM is somewhat pro-social and has an alpha of 0.5 (which corresponds to an SVO angle[1] of about 27B°), then we would expect DM to only cooperate if he is at least 67% sure that the other player will also cooperate.

This trust threshold[2] obviously depends on the payoff structure of a particular PC game. A trust threshold (like that shown in Figure 3 and Figure 4) can be found with Eq.(6) and Eq.(7) given a PC game with payoffs $CBAD$, and a joint utility function similar to equation:

$$\beta_1 = \frac{\alpha_2 A - \alpha_2 B - A - D}{\alpha_2(A - B + C - D) + A + B - C - D} \tag{6}$$

$$\beta_2 = \frac{\alpha_1 A - \alpha_1 C + A - D}{\alpha_1(A + B - C - D) + A - B + C - D} \tag{7}$$

Next, we consider the 36 PC games listed in Table 1. The model predictions for all these different PC instances are shown in Figure 5 and Figure 6. PC games with the same K-index cooperative characteristics are in the same group. There are a few things to note from this visual explanation. First, as the K-index increases, the size of the cooperative region (light shading) generally increases as well. This makes an intuitive sense, as it shows which those PC games that require less SVO or less positive expectations to demonstrate trust are also those where we expect more cooperation. There is general agreement between K and the model for the severity of a particular PC. However, the relationship between K and the position and curvature of the trust threshold is not completely linear. The model distinguishes between PC games that share the same K-index, and thus provides a more granular understanding of the severity of PC games and better predictions of when cooperation will emerge.

Figure 5 and Figure 6 show trust thresholds for 36 PC games. As these figures, each subplot has an x-axis corresponding to SVO (α_2), and a y-axis corresponding to positive expectation (β_1). Rapoport's k-index is shown above, and the payoff structure ($CBAD$) that defines each specific PC game is shown above each subgraph. The light zone matches the expected cooperation.

Figure 7 shows trust thresholds for 20 PC games. In these Figures, each subplot has an x-axis corresponding to SVO (α_1), and a y-axis corresponding to positive expectation (β_2). Rapoport's k-index is shown above, and the payoff structure ($CBAD$) that defines each specific PC game is shown above each subgraph. The light zone matches the expected cooperation. This figure only shows a negative slope of the trust threshold, and there is no counter-intuitive situation, as the SVO (α_1) increases, the positive expectation (β_2) begins to decrease.

---

[1] α values can be translated into SVO angles by the following equation: $tan(\alpha) = SVO^\circ$ and can be translated back the other way via $SVO^\circ = arctan(\alpha)$. For values of α between 0 and 1, this is close to a linear transformation. For more information regarding computations of SVO angles, see Murphy, Ackermann, and Handgraaf (2011). Traditionally, SVO scores are reported as angles, but rescalingthem produces values that are more readily interpretable.

[2] The trust threshold generally refers to DM's actual credit recognition of another player, and the threshold is mainly determined by DM's subjective consciousness.

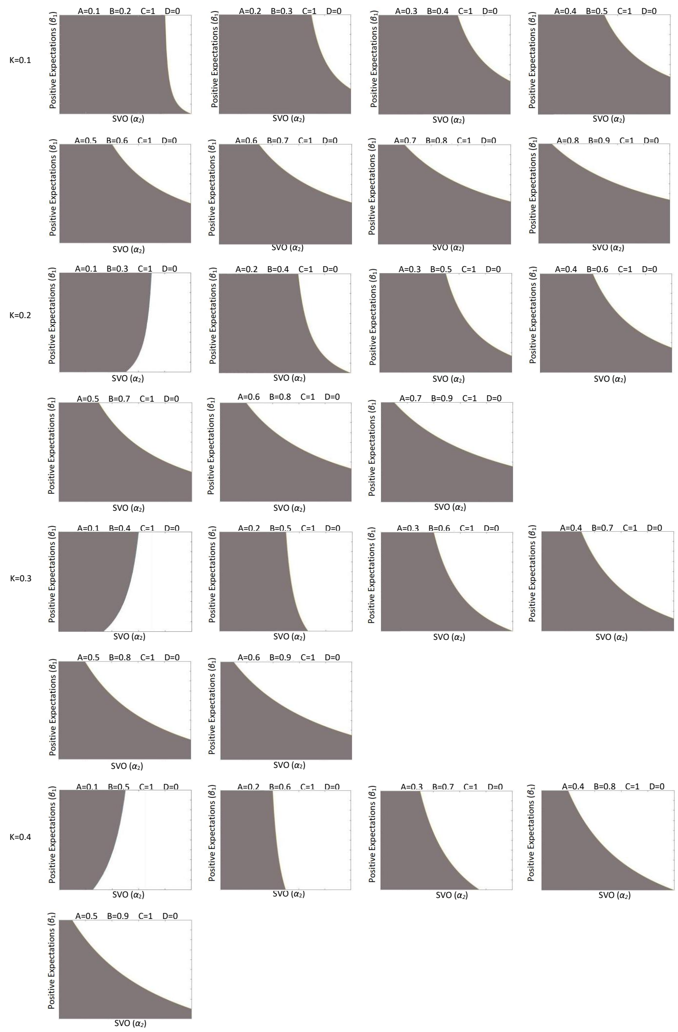


Fig. 5. Figures of trust thresholds for PC games (for Player 1)

We can derive a new cooperation index from the model's predictions. The CoopArea[3] (bright area) is another indicator of the severity of the PC game.

[3]In this paper, the cooperation area refers to the area of the brightly colored area in the figure.

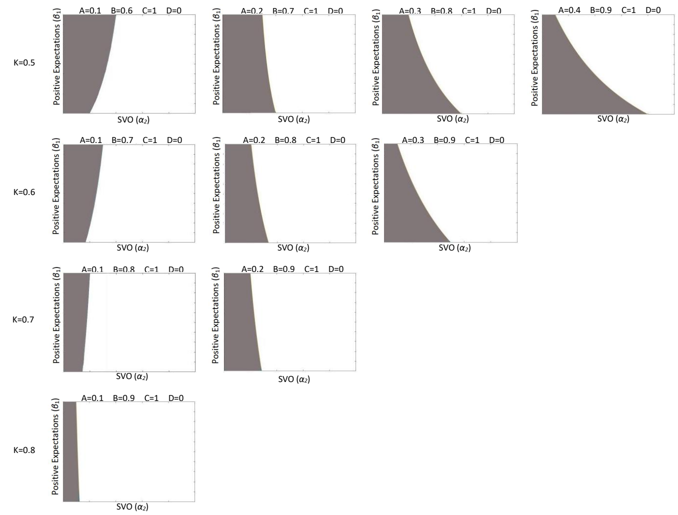


Fig. 6. Graphs of trust thresholds for PC games (for Player 1)

Alternatively, one can also determine the value of α that is minimally sufficient to induce cooperation on the basis of the principle of insufficient reason, assuming that $\beta_1 = \beta_2 = 0.5$. Such an exponent (α_{crit}[4]) can be understood as the answer to the question:вЂњ How much SVO does a DM need to prove cooperation, since DM doesnot know (e.g., before a unified) what other players need to choose?вЂќ. This critical SVO level[5] can be found for a PC game with the following equation:

$$\alpha_{1-crit} = -\frac{A + B - C - D}{A - B - C + D} \tag{8}$$

$$\alpha_{2-crit} = -\frac{A - B + C - D}{A - B - C + D} \tag{9}$$

The α_{crit} values of different PC games, the corresponding cooperation areas, and the corresponding ranking correlation K-index are shown in the following figure, which also provides a visualization of the relationship between these three cooperation indices in the form of a bivariate scatter plot.

In addition, another well-known indicator can also be considered as the cooperation index, the "Price of Anarchyээ (PoA; see Mak and Rapoport, 2013). In the context of PC Game, PoA is simply the ratio of $\frac{B+C}{2A}$. Clearly, the metrics extracted from the model predictions are closely related to the K-index of cooperation, which may provide some insight into why the K-index can predict cooperation rates in PC Games in the first place, i.e. because it captures some interaction between

[4] α_{crit} refers to α critical value (also called threshold), that is, the critical point between two different states of things, is the watershed where the state of things jumps.

[5] This critical value uses the joint utility function as stated in Eq. (1). A different utility function would yield a different formula for α_{crit}.

preferences, beliefs, and payoffs the psychology behind it. In fact, the K-index corresponds almost perfectly (negatively) to the minimum degree of prosociality required for trusting and cooperative action under the principle of insufficient reason. But the K-index is not as sensitive as the CoopArea index because it only considers one of the three components of the interaction (i.e. payoff), whereas the cooperation area is influenced by three factors (i.e. preference, belief, and payoff).

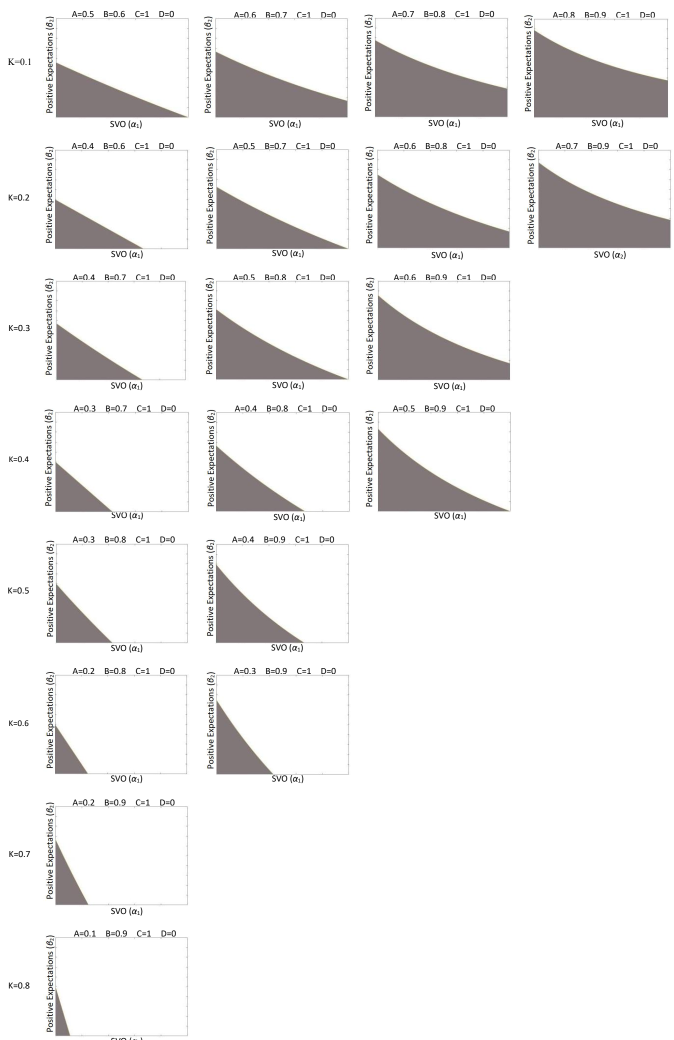
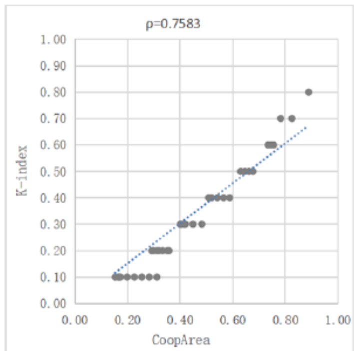
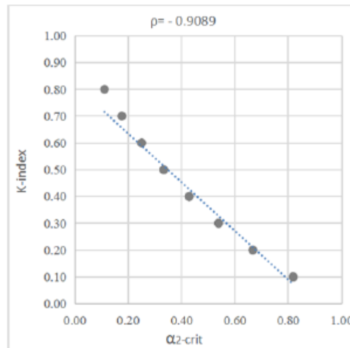


Fig. 7. Figures of trust thresholds for PC games (for Player 2)

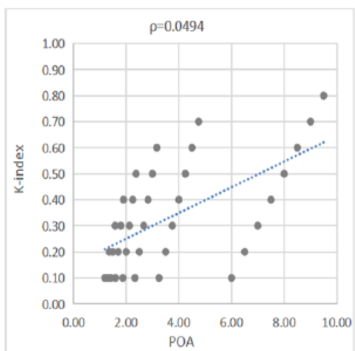
Next, we will use the data in Table 1, 2 (see Figure 8 and Figure 9) to fit the linear relationship plots among the cooperation indicators such as K-index, PoA, CoopArea, and α_{crit}, and then generate trend lines, and finally predict their effects on the cooperation rate based on the slope of the trend line ρ[6], which is negatively correlated if $\rho < 0$ and positively correlated if $\rho > 0$.



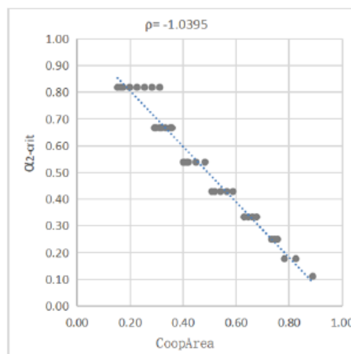
(a) CoopArea and K-index



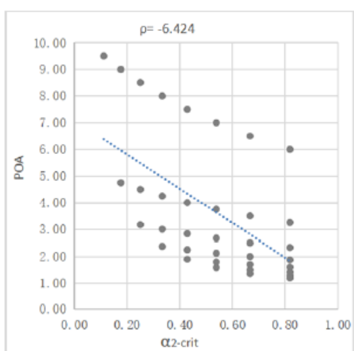
(b) α_{2-crit} and K-index



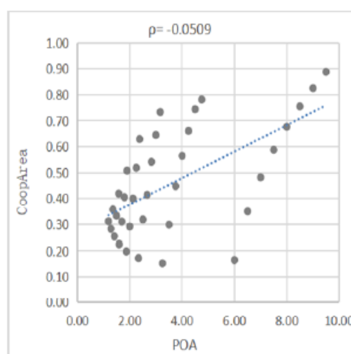
(c) PoA and K-index



(d) CoopArea and α_{2-crit}



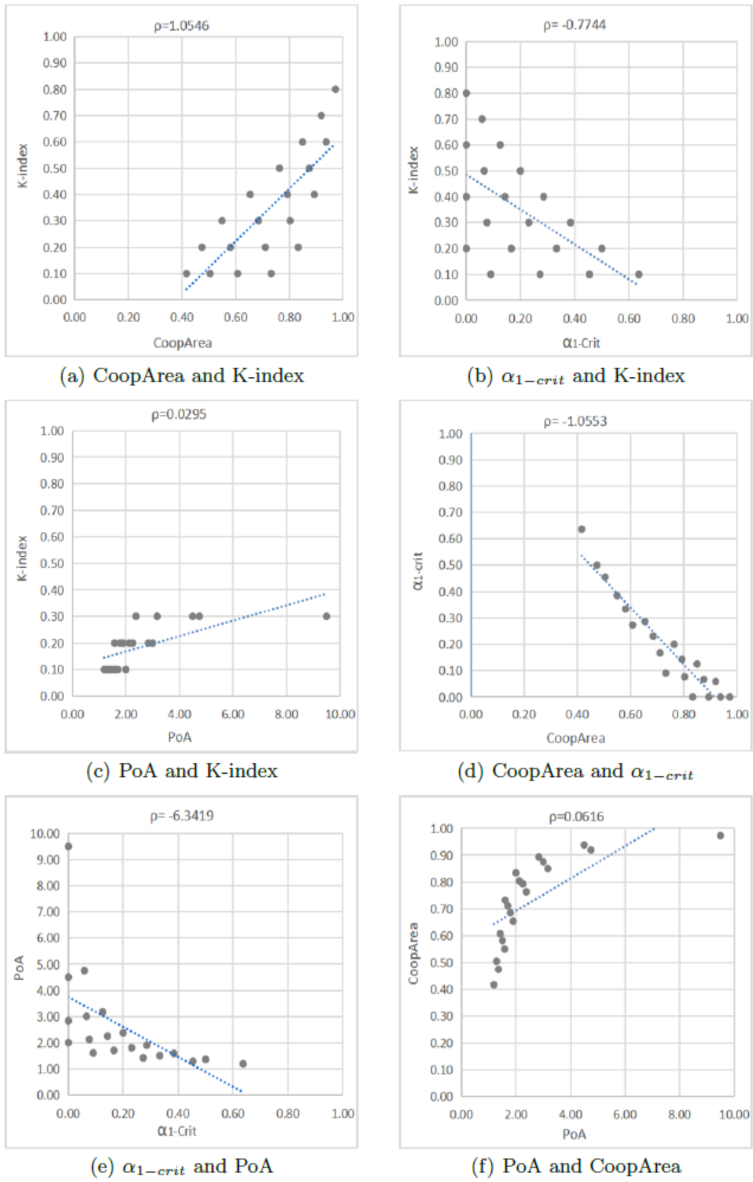
(e) α_{2-crit} and PoA



(f) PoA and CoopArea

Fig. 8. The relationship of Cooperation indexes (for Player 1)

[6]In the above graphs, ρ is the slope of the trend line of each relationship graph.



(a) CoopArea and K-index (b) α_{1-crit} and K-index

(c) PoA and K-index (d) CoopArea and α_{1-crit}

(e) α_{1-crit} and PoA (f) PoA and CoopArea

Fig. 9. The relationship of Cooperation indexes (for Player 2)

From Figure 8 and Figure 9, we can see the following rules:
(a) With the increase of CoopArea, the K-index will also increase;
(b) With the increase of α_{crit}, the K-index will decrease;
(c) With the increase of PoA, the K-index also increases;
(d) With the increase of CoopArea, the α_{crit} decreases;
(e) With the increase of α_{crit}, the PoA will decrease;
(f) With the increase of PoA, the CoopArea will increase.

To sum up, α_{crit} is inversely proportional to other indices, and other indices are directly proportional to each other. we can obtain the cooperation indicators that can promote the cooperation rate, including CoopArea, K-index, and PoA.

5. Conclusion

Using the standard 2Γ—2 PC game as the most prominent paradigm, we address several issues in describing trust-based cooperative processes in social dilemmas. Our main findings can be summarized as follows.

First. Trust-based cooperation in social dilemmas, such as PC games, is reasonable within the framework of subjective expected utility because DM has: (a) sufficient prosocial preferences; (b) sufficiently aggressive to expect that other players will choose to cooperate; (c) Potential payoffs form a sufficiently insignificant game. This approach accommodates the heterogeneity of decision makers in preferences and beliefs and can constructively explain the observed individual differential choice behavior, while still accommodating the human economy as a literal corner case ($\alpha = \beta = 0$).

Second. A "trust threshold" can be derived in the preference-belief (α, β) space, which defines which combinations of SVO and positive expectations that trust-based cooperation is rationalizable and is the best response. In PC games, these thresholds are related to the K-index, but not exactly the same. Furthermore, the preference - belief space (Figure 3 and Figure 4) is a useful framework for considering individual differences, especially in cases where DM is close to cooperation but not quite above the threshold.

Third. A counterintuitive result of this model is that the slope of the trust threshold line is sometimes positive. This means that for some PC games, the model predicts that DMs will be more likely to cooperate because of lower expectations for other players. This is the result of a simple joint utility model from equations. To ensure that only negatively sloped trust thresholds occur, a different social utility model needs to be implemented. For example a joint utility function with a contingent component could be used (e.g., DM's social preference is positive if and only if the other player chooses H (or h), and zero otherwise). With the use of more sophisticated contingent social preference models, trust thresholds have changed in shape and have different properties, some of which may be desirable from a descriptive perspective.

Finally. Our findings support some caveats against relying too heavily on behavioral theoretical predictions in strategic situations, and these simulation studies come from investigations of single PC games with specific payoff structures. A simple decision model can make significantly different predictions about behavior in the same game due to the interaction between the social utility model and payoff. Often, researchers do not report comprehensive instances of a particular type of game, so they may over-generalize results to other contexts when predictions may not even apply to different instances of PC games with different payoffs.

Acknowledgement

Thanks to Prof. Elena Parilina for her valuable feedback on our discussions and earlier versions of this paper.

References

Andreoni, J. and Miller, J. H. (2002). *Giving according to GARP: An experimental test of the consistency of preferences for altruism.* Econometrica, **70(2)**, 737–753.

Au, W. and Kwong, J. (2004). *Measurements and effects of social-value orientation in social dilemmas: A review.* In: R. Suleiman, D. Budescu, I. Fischer, & D. Messick (eds.), Contemporary psychological research on social dilemmas, 71–98. New York: Cambridge University Press.

Axelrod, R. (1967). *Conflict of interest: An axiomatic approach.* Journal of Conflict Resolution, **11**, 87–99.

Balliet, D. and Van Lange, P. A. (2013). *Trust, conflict, and cooperation: a meta-analysis.* Psychological Bulletin., **139(5)**, 1090–1112.

Berg, J., Dickhaut, J. and McCabe, K. (1995). *Trust, reciprocity, and social history.* Games and Economic Behavior, **10(1)**, 122–142.

Glöckner, A. and Hilbig, B. E. (2012). *Risk is relative: Risk aversion yields cooperation rather than defection in cooperation-friendly environments.* Psychonomic Bulletin & Review, **19(3)**, 546–553.

Kreps, D. M., Milgrom, P., Roberts, J. and Wilson, R. (1982). *Rational cooperation in the finitely repeated prisoners' dilemma.* Journal of Economic Theory, **27(2)**, 245–252.

Mak, V. and Rapoport, A. (2013). *The price of anarchy in social dilemmas: Traditional research paradigms and new network applications.* Organizational Behavior and Human Decision Processes, **120**, 142–153.

Murphy, R. O. and Ackermann, K. A. (2014). *Social value orientation: Theoretical and measurement issues in the study of social preferences.* Personality and Social Psychology Review, **18(1)**, 13–41.

Murphy, R. O., Ackermann, K. A. and Handgraaf, M. J. J. (2011). *Measuring social value orientation.* Judgment and Decision Making, **6(8)**, 771–781.

Rapoport, A. (1967). *A note on the index of cooperation for Prisoner's Dilemma.* Journal of Conflict Resolution, **11(1)**, 100–103.

Rapoport, A. and Chammah, A. M. (1965). *Prisoner's dilemma: A study in conflict and cooperation.* Ann Arbor: University of Michigan Press.

Roth, A. E. and Murnighan, J. K. (1978). *Equilibrium behavior and repeated play of the prisoner's dilemma.* Journal of Mathematical Psychology, **17**, 189–198.

Rousseau, D., Sitkin, S., Burt, R. and Camerer, C. (1998). *Not so different after all: A cross-discipline view of trust.* Academy of Management Review, **23(3)**, 393–404.

Steele, M. and Tedeschi, J. (1967). *Matrix indices and strategy choices in mixed motive games.* Journal of Conflict Resolution, **11(2)**, 198–205.

Van Lange, P. A. M. (1999). *The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation.* Journal of Personality and Social Psychology, **77(2)**, 337–349.

Vlaev, I. and Chater, N. (2006). *Game relativity: How context influences strategic decision making.* Journal of Experimental Psychology: Learning, Memory, and Cognition, **32(1)**, 131–149.