

Lans Bovenberg  
*Tilburg University*

Bas van Os  
*Erasmus University Rotterdam*

*Abstract.* The well-known prisoner's dilemma shows why cooperation may be difficult to achieve even when cooperation would create value for all partners. In particular, material interests conflict because reciprocity cannot be enforced, while agents give priority to their material self-interest. This paper introduces a model of human motivation based on faith and love, and shows how these motivations transform the prisoner's dilemma into a stag hunt game with a cooperative and a non-cooperative equilibrium.

*Keywords:* prisoner's dilemma, strong reciprocity, relational goods, faith, love

## 1. Introduction

Even though it can create value for all, cooperation as mutual service is often difficult to achieve because the lack of direct reciprocity between giving and receiving creates conflicts of material interest between decisionmakers. This paper discusses how a slightly more complex model of human motivation than material self-interest may help to explain why cooperation is often possible, even in a setting in which reciprocity between giving and receiving material goods is difficult to enforce. Section 2 introduces this model of human motivation for simultaneous games in which agents have only the present time to consider.<sup>1</sup>

Section 3 outlines the well-known prisoner's dilemma. In this game, individual material interests conflict with the collective interest because of lack of reciprocity between giving and receiving. Indeed, in the case of complex cooperation, transaction costs make it hard to enforce reciprocity through complete contracts. By giving priority to their material self-interest, individuals then destroy value by harming rather than serving each other.

---

<sup>1</sup> Bovenberg & van Os (2025) expand this model to sequential games in which agents account for the strategic effects of their choices on the behavior of others in the future.

In real life, however, agents often choose to co-operate in the prisoner's dilemma, even if it is a one-shot game or a finitely repeated game (Fehr and Fischbacher, 2003).<sup>2</sup> Section 4 shows how the model of human motivation outlined in section 2 turns the prisoner's dilemma into a stag hunt game with both a cooperative and a non-cooperative equilibrium. In particular, positive feedback between trust and cooperation causes both trust and fear to be self-fulfilling. In this way, subjective expectations have direct consequences for the creation of material value.

## **2. Human motivation**

### *Homo amans*

Patrick Nullens and Jermo van Nes (2022) present the 'loving human' (*homo amans*) as an alternative to the model of the *homo oeconomicus*, which economists employ to explain human behavior in perfect markets in which individual decisions to trade do not affect others. In relational settings in which individual choices impact the welfare of others, however, the *homo amans* may provide a better descriptor of human behavior. In the language of economics, neighborly love is understood as the mental internalization of external effects. In this view, rational agents are intrinsically motivated to attach value to the effects of their decisions on the welfare of neighbors. Neighbors are defined as those people whose welfare is affected by the agent's decisions. Accordingly, neighbors have a stake in the agent's decisions and can therefore be called stakeholders. Mentally internalizing external effects on one's neighbor differs from enlightened material self-love when an agent takes into account the effects on one's neighbor only for strategic reasons, i.e. to the extent that this yields (future) benefits for the agent.

### *Two types of love*

Borrowing from Lewis (1960), we adopt Greek words for love to distinguish between various motivations. For simultaneous games, we distinguish between two types of love:

- *Eros*, or self-love, in terms of direct material self-interest.
- *Philia*, or conditional neighborly love. It is oriented towards the friends and the groups with which the agent identifies (as opposed to people to whom the agent is indifferent or regards as enemies or traitors).

---

<sup>2</sup> Individuals who care only about their material self-interest may cooperate in an infinitely repeated prisoner's dilemma because of strategic effects of their behavior on future choices of the other player. In particular, individuals realize that their present cooperation will be rewarded in future games through continued cooperation of others. However, self-interested players do not cooperate in a one-shot prisoner's dilemma or in a finitely repeated prisoner's dilemma with a known end game. In these settings, the motivations outlined in section 2 can explain cooperation.

*Material and Relational Goods*

The two types of love relate to two types of goods that the agent can derive from a decision to co-operate:

- *Eros*: like all animals, humans are driven by their bodily states to pursue food, shelter, and sex. The satisfaction associated with bodily states can be viewed as material goods that the agent enjoys.
- *Philia*: like all social animals, humans are more successful in meeting their material needs if they co-operate with others. Hence, they have developed a need for belonging and therefore experience feelings of strong reciprocity. These feelings motivate people to reward those who cooperate or are expected to cooperate, even if it requires giving up material goods (Bowles 2010). The emotional satisfaction associated with relational status can be viewed as relational goods that the agent enjoys (Bruni and Stanca 2008).

*Neighborly love is self-love too*

When asked about the greatest commandment, Jesus of Nazareth answered, first, to love God and, second, to love your neighbor like yourself.<sup>3</sup> Rather than denigrating self-love, Jesus uses it as the standard for loving neighbors. Loving one's neighbor is not a burden to be carried at the expense of one's own joy and well-being, but rather a profound understanding of human nature: people enjoy working towards the well-being of the people and groups they love.

*Trusting and loving God is the foundation of neighborly love*

The mere suspicion that the other is an adversary may destroy our natural inclination to cooperate. The question is thus whether the agent trusts the other to reciprocate. Here 'faith,' which in the Greek of the New Testament also means trust or loyalty, can provide a more secure base. Jesus saw trust (faith) in the fatherly love of God as the foundation of his ethics. The fatherhood of God makes humans part of the same household. Even though someone may not always reciprocate that love, the love of their heavenly father acts as a collateral.<sup>4</sup> Loving God above all by trusting God therefore always produces a relational good when rational agents love their neighbors. These relational goods are also produced in cultures that nurture social capital and positively affirm cooperative behavior among their members. These goods can thus be related to social capital as the trust and the respect in a society or household.

---

<sup>3</sup> Matthew 22: 36-40.

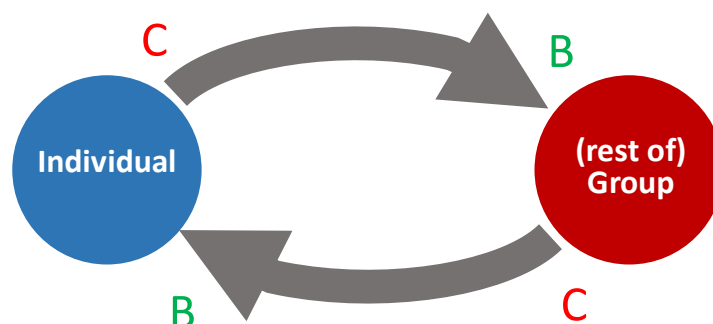
<sup>4</sup> Matthew 6: 30-4 and Romans 8:37-8.

### 3. Cooperation: valuable but difficult to organize

#### *Abundance and scarcity*

Economics can be viewed as the science that studies the creation of value for all, through cooperation, where cooperation is defined as mutual service. The normative law of economics is thus mutual benefit (Sugden 2018): creating winners without any losers. Economics is thus about abundance, in that cooperation can create surplus for all. At the same time, economics is about scarcity: if you want to obtain a valuable good, you have to give up another valuable good.<sup>5</sup> Choice is thus inevitable: no gain without pain. Reciprocity (for a scarce good you have to give up another valuable good) on account of scarcity as a natural, biological law translates into reciprocity as a normative, social law governing cooperation.<sup>6</sup> In particular, reciprocity between giving and receiving ensures that all partners share in the gain from cooperation, and thus face an incentive to participate in and contribute to the cooperation. Unlocking the potential of cooperation in creating value thus involves sacrifices of all participating members in order to ensure that all of these members gain.

Figure 1. Cooperation as exchange.



#### *Bilateral trade*

Figure 1 illustrates how economics can be about abundance (i.e. creating surplus for all) and scarcity (the requirement to sacrifice goods) at the same time, in the case of exchange as simple, symmetric bilateral trade. If both partners give up  $C$ , they both will gain  $B$ . What they give up, they both value less than what they receive (i.e.  $C < B$ ). Hence, they both gain a surplus, namely  $B - C > 0$ . To give a specific example of such a situation, consider two youngsters, Ann and Ben. Ann is good in mathematics and Ben in English. If Ann spends one hour of her time teaching mathematics to Ben, Ben will save one and a half hours of study time. Similarly, Ann will save one and a half hours studying English if Ben is willing to devote one hour to helping

<sup>5</sup> The word 'good' is defined here as anything that is good in the sense the people derive welfare from it. Hence, a good does not have to be material but can be an immaterial service or experience.

<sup>6</sup> This is in fact related to the well-known categorical imperative as formulated by Immanuel Kant.

her. Hence, the exchange of heterogeneous goods is symmetric. By helping each other, they will both give up one hour of their time for teaching the subject they are good at to gain one and a half hours study time in the subject they need help in. Hence, they both gain a half hour (one and a half minus one hour). The cost  $C$  is one hour, and the benefits one and a half hours, so that the return on the investment  $(B-C)/C$  amounts to 50%.

#### *Public goods*

Another example is cooperation in producing a public good such as a clean shared home. If all members of a household give up an additional 10 minutes during the week by putting their litter in the waste baskets, they will each save 50 minutes at the end of the week, in terms of having to clean their home. In this case,  $B = 50$  minutes and  $C = 10$  minutes, so that the gain from cooperation (i.e. not littering) is  $B-C = 40$  minutes for each household member. In this case, the potential return on the investment is  $(B-C)/C = (50-10)/10 = 400\%$ .

#### *No conflicts of interest in transactions*

In a transactional environment of so-called direct exchange, people can receive something from others only if they at the same time give something back (i.e. 'pay') to those who serve them. Moreover, if they serve others, they are immediately rewarded through a payment by those whom they serve. As a result of this reciprocity between giving and receiving, parties do not face a conflict of interest between serving themselves and serving others. Indeed, they can serve themselves only if they serve others. Cooperation by serving others is enlightened material self-interest.

#### *Separating receiving from giving causes conflicts of interests*

Cooperation becomes more difficult if people do not engage in direct exchange between giving and receiving. In these more relational environments, the complexity of cooperation can originate in the goods exchanged and/or the number of partners involved. Without a direct relationship between giving and receiving on an individual level, individuals face a conflict of material interest between their individual interest and the collective interest of the group as a whole. Hence, cooperation fails if the individuals give priority to their individual interests.

#### *Conflicts of interest destroy value if people care only for material self-interest*

The so-called prisoner's dilemma captures the fundamental challenge of creating cooperation in relational, more complex environments. In these settings, high transaction costs preclude complete contracts that could have ensured reciprocity between giving and receiving at each point in time. The game shows how a lack of direct reciprocity between giving and receiving results in conflicts between, on the one hand, the interest of the decisionmakers (so-called agents) and, on the other, the interests of the stakeholders that are affected by the decisions of these decisionmakers (so-called principals, or neighbors). Moreover, it describes how these conflicts of interest between agents and principals destroy value for all concerned in the event

that the decisionmakers care only about their own material self-interest and thus do not shy away from opportunistically benefitting themselves at the expense of others—even if they promised earlier to protect their interests.<sup>7</sup>

**Table 1. Prisoner's dilemma.**

Anna / Ben	Cooperate	Defect
Cooperate	(B-C, B-C)	(-C, B)
Defect	(B, -C)	(0, 0)

*The pay-offs in a prisoner's dilemma*

Table 1 shows the material pay-off matrix of the prisoner's dilemma. The material pay-offs<sup>8</sup> are related as follows:  $B > (B-C) > 0 > -C$ . Cooperation is thus valuable because the joint pay-off of both players together is the highest in the cooperative equilibrium and the lowest in the non-cooperative equilibrium:  $2(B-C) > B-C > 0$ . The prisoner's dilemma is not a zero-sum game: cooperation is efficient in the sense that it creates value for the household as a whole. Moreover, both players are better off in the cooperative situation than in the non-cooperative situation (because  $B-C > 0$ ). The cooperative situation thus Pareto dominates the non-cooperative situation.

*Dominant strategy and non-cooperative equilibrium*

Even though cooperation benefits both players, the players end up in the non-cooperative situation if they pursue only their material self-interest and take the decisions of other players as given. Indeed, the dominant strategy of both players is then to play non-cooperatively. Hence, the dominant-strategy equilibrium is the non-cooperative equilibrium in which both players earn a pay-off of 0. In particular, starting from the cooperative situation in which both players cooperate, a player can improve his individual pay-off from  $B-C$  to  $B$  by defecting. This

<sup>7</sup> Hence, a communication phase (in which players can make agreements before the players actually make their decisions) in the prisoner's dilemma does not succeed in creating cooperation unless such agreements can be enforced by an independent party.

<sup>8</sup> The pay-offs are material in the sense that the goods involved are tradable and can thus be exchanged for other goods (or for money) on markets. In addition to material (or biological) goods, section 4 distinguishes also relational (or social) goods. The material pay-offs in Table 1 can be viewed as narrow welfare because the pay-offs do not include the relational goods that are considered in section 4, in general, and in Table 3, in particular.

defection from cooperation imposes external costs of  $B$  on the other player, so that the improvement of the individual pay-off of  $C$  goes at the expense of a loss in the joint pay-off for the household as a whole of  $(B-C > 0)$ . The other player can then limit his losses by also defecting. This improves his pay-off, from  $-C < 0$  to  $0$ , but imposes external costs of  $B$  on his partner and a loss of  $(B-C)$  on the household as a whole. For both players, the gain from betraying the other player ( $C$ ) is smaller than the loss of being betrayed ( $B$ ), so that both players lose  $(B-C)$  in the non-cooperative equilibrium compared to the situation in which they would both cooperate.

#### *Cooperation fails because of material self-interest and lack of reciprocity*

The reason for this perverse result is twofold. First, the individual material interests conflict with the collective material interest because the players can take without returning anything: reciprocity between giving and receiving is not enforced.<sup>9</sup> Second, if material individual and collective interest conflict, the players who care only about their material self-interest give priority to their individual interests over and above the collective interest.

#### *Conflict of interest: Collective return on individual investment*

The conflict between the individual and collective material interests is captured by the parameter  $R \equiv \frac{B-C}{C} > 0$ . Here  $(B-C)$  stands for the collective ('we') interest (i.e. the benefits for the group as a whole as a result of the decision to cooperate) and  $C$  for the individual ('I') material interest (i.e. the costs of cooperation for the individual who makes the decision). This ratio between net benefits  $(B-C)$  and costs  $(C)$  is the collective ('we') return of the individual investment. Since the return is positive, the investment is efficient from the household (i.e. collective) point of view: the benefits  $(B)$  exceed the costs  $(C)$ . Since the material costs are for the individual, and the material benefits for the rest of the group, there is no reciprocity between individual costs and individual benefits. The absence of strategic effects of giving on the behavior of others leads to a conflict of interest.

#### *Multilateral exchange with public goods*

The prisoner's dilemma has in fact the same incentive structure as a public goods game. In both games, the collective (material) interest clashes with the individual (material) interest because of lack of reciprocity. In the public goods game, agents can benefit from the public good without contributing to it, just as in the prisoner's dilemma an individual can receive a good from his partner without giving anything back to that party. In both games, direct reciprocity on an individual level between giving and receiving is lacking. The parameter  $R$  for

---

<sup>9</sup> With direct exchange, collective and individual material interests would coincide because the individual could reap the gain from the gift of the other party ( $B$ ) only if that individual is willing to give up  $C$ . In other words, there are strategic effects: an individual giving up something valuable to others leads others to sacrifice something valuable for his or her benefit. Reciprocity at an individual level thus would be enforced and the asymmetric, free-rider and hold-up cases in Table 1 (in which one party gives without receiving anything) would be absent.

the ratio between the collective benefits and the individual costs of an individual contribution to the public good captures the key incentive problem (i.e. the conflict between the individual material interest [involving a negative pay-off  $C$ ] and collective interests [involving a positive pay-off  $B-C$ ]).<sup>10</sup> Table 2 provides the pay-off matrix of the prisoner's dilemma in terms of this single parameter  $R$ .

**Table 2. Prisoner's dilemma in terms of parameter for conflict of interest ( $R>0$ ).**

Anna / Ben	Cooperate	Defect
Cooperate	(R, R)	(-1, 1+R)
Defect	(1+R, -1)	(0, 0)

#### 4. Social preference of strong reciprocity in simultaneous games

##### *Social preferences*

We speak of social (or other-regarding) preferences if individuals value not only their own material consumption but also that of others (Dhami 2019). We can write social preferences in a simple linearized form as

$$s_i = w_i + \sum_{\substack{j=1 \\ j \neq i}}^N (\alpha_{ij} * w_j). \quad (1)$$

Here,  $N$  denotes the number of individuals that interact,  $s_i$  is the overall welfare (or pay-off<sup>11</sup>)

<sup>10</sup> In a public goods game, the costs  $C$  should be interpreted as the net costs (i.e. the gross costs minus the benefits that the individual derives from their own contribution). In a symmetric public goods game with  $N$  players, the benefit that an individual derives from their own (marginal) contribution is  $B/N$ . If  $N$  becomes very large, the benefits an individual collects from their own contribution thus become negligible so that the net costs  $C$  equal the gross costs of the contribution.

<sup>11</sup> All pay-offs are measured in terms of the same unit of account, which can be money (e.g. euros). However, these pay-offs are not necessarily material pay-offs, in the sense that they can be transformed on the market into money as a medium of exchange. Indeed, the pay-offs of individual  $I$  involving the moral sentiments for the material pay-offs of others ( $j \neq i$ ) are non-pecuniary.

of individual  $i = 1, 2, \dots, N$ , and  $w_i$  is the material<sup>12</sup> pay-off of individual  $i = 1, 2, \dots, N$ . The conception of welfare is now broader than material welfare alone. People are motivated by love for not only their own material goods but also the material goods of others. The ‘love’ parameter measuring the moral sentiments of player  $i$  for player  $j$  is  $\alpha_{ij}$ .<sup>13</sup>

### *Broader welfare*

The welfare of people with social preferences is broader than just their own material consumption. Cooperation as serving the common interest is not only an instrument for a person’s own material consumption  $w_i$  but also contributes to the immaterial well-being of person  $i$  by meeting the material needs of others for whom  $i$  has positive moral sentiments so that  $\alpha_{ij} > 0$  for  $i \neq j$ . Indeed, people derive welfare not only from receiving material goods but also from giving material goods to a group with which they identify themselves (see section 2).

### *Neighborly love as internalization of external effects*

We define neighborly love as the mental internalization of these external effects—in the sense that you include the external effects of your actions on others (i.e. neighbors) in your own welfare calculations. Hence, if person  $i$  loves person  $j$  as himself, person  $i$  includes the material welfare of person  $j$  in his own welfare calculus with the same weight, so that  $\alpha_{ij} = 1$  in equation (1). If a person mentally internalizes the external effects on others in a group, that person considers the collective welfare of a group as their own. In this case of so-called we-rationality (Smerilli 2007), conflicts between the collective (‘we’) interest<sup>14</sup> and the broad individual (‘I’) interest are absent. Hence, even in the absence of strategic effects due to direct reciprocity (between giving and receiving in a material sense), people’s decisions are efficient in the sense that they maximize the welfare of the group (or household) as a whole.

<sup>12</sup> Material pay-offs involve the consumption of material goods by the individual. The love of an individual for the pay-offs that others derive from material goods results in relational goods for that individual. These non-pecuniary pay-offs involve the moral sentiments identified by Adam Smith as follows: “How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it.” Note that the self-interest or selflove can be defined in two alternative ways. In particular, self-interest can be defined narrowly by confining it to the material pay-off  $w_i$ . However, selflove may be defined also more broadly as  $s_i$ . In that case, it includes also the non-material, non-pecuniary components  $\alpha_{ij} * w_j$  ( $j \neq i$ ). This paper defines self-love (or welfare) more broadly.

<sup>13</sup> If this parameter is negative, player  $i$  is spiteful with respect to player  $j$ .

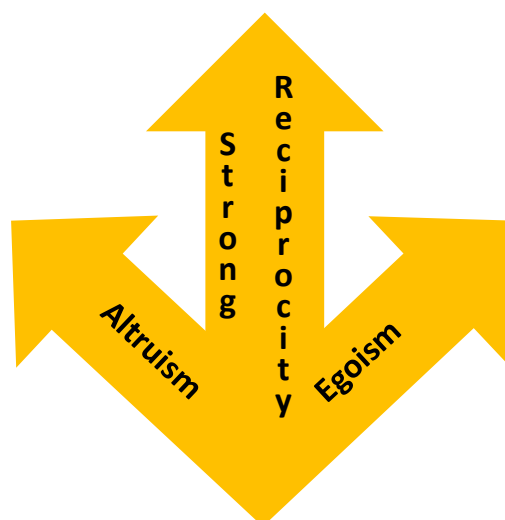
<sup>14</sup> We measure group welfare as the sum of the welfare levels (measured in the same unit of account) of the individuals making up the group. Measuring the ‘common interest’ of a group this way can be done if welfare is quasi-linear in one of the tradable (or transferable) goods and if group members can transfer tradable goods among themselves. These two conditions yield transferable utility and ensure that the (efficient) allocation of resources is independent of the distribution of property rights (Medema, 2020).

*Individual behavior*

*Social preference of strong reciprocity*

Most people are not unconditionally altruistic, in the sense that they always mentally internalize the external effects of their behavior on others. In fact, they mentally internalize external effects on the material welfare of others only if these others act cooperatively toward them or are expected to act cooperatively. In the setting of a simultaneous game, the 'love' (or 'moral sentiment') parameter of player  $i$  for a partner  $\alpha_{ij}$  is 1 if this partner  $j$  cooperates and 0 if the partner  $j$  defects.<sup>15</sup> The literature on social preferences calls this type of social preference 'strong reciprocity' (Bowles 2004).<sup>16</sup> This model of social preferences is popular in experimental economics (Fehr and Fischbacher 2003), evolutionary biology (Gintis 2000 and Bowles 2016) and evolutionary psychology (Seabright 2010). A person with these social values is willing to reward people who are good to him, even if these rewards impose material costs on him. These social preferences correspond to the *philia* love described in section 2.

Figure 2. Strong reciprocity



*Strong reciprocity as a third way between selfishness and altruism*

Strong reciprocity as a social preference can be viewed as a third way between a naive, optimistic view of humanity and a skeptical, pessimistic stance. On the right-hand side of Figure 2 is a selfish player who is unconditionally non-cooperative in the prisoner's dilemma (or the public goods game) in which individual material interests (the 'I interest') conflict with the collective interest (the 'we interest'). Hence, even if others contribute to a public good, this

<sup>15</sup> They may even value the pay-off of their partners negatively if these partners defect. We may then speak about 'negative intrinsic reciprocity'.

<sup>16</sup> The moral sentiment parameter thus depends on the actions of others rather than on their intentions. See Dhami (2019) for models in which intentions play a role, such as Rabin (1993). For a critical review of preferences based on the unobservable intentions of others rather than the observable actions of others, see Sugden (2018).

player does not. This person is thus a free-rider who does not mind acting like a parasite on those who help him. On the left-hand side of figure 2 is the saint who is altruistic with respect to all neighbors, even those who are harming that person or the group with which that person identifies. This person is thus unconditionally cooperative in the event of a conflict between individual and collective material interests. A strongly reciprocal person (the one in the middle of figure 2) is conditionally cooperative. This person internalizes external effects on others if these others (are expected to) act cooperatively towards that person. One thus cooperates as long as one expects others to do the same; in other words: for as long as one trusts others.

**Table 3. Prisoner's dilemma in material pay-offs with intrinsic reciprocity.**

Anna / Ben	Cooperate	Defect
Cooperate	$(R+R, R+R)$	$(-1, 1+R-1)$
Defect	$(1+R-1, -1)$	$(0, 0)$

*Broad welfare: pay-offs with strong reciprocity*

Table 3 shows the pay-off matrix if the players feature social preferences of strong reciprocity in case the material pay-offs are the same as in the prisoner's dilemma (or public goods game) from Table 2. The bold pay-offs are the additional pay-offs on account of moral sentiments. In particular, if their counter-party cooperates, the players internalize the external effects of their choices so that they then consider the collective ('we') interest rather than the individual ('I') material interest. In other words, due to their moral sentiments, players consider a concept of welfare that is broader than just their individual material welfare.

*Individual behavior: conformist tit-for-tat strategy rather than free-riding*

If they exhibit strong reciprocity, the players no longer feature a dominant strategy. They in fact follow a kind of tit-for-tat strategy in which they conform to the strategy of the other player. The difference with the prisoner's dilemma is thus that they no longer defect but cooperate if the other player cooperates. In other words, players featuring social preferences

of strong reciprocity are not free-riders: they dislike betraying a person who (they expect) is good to them.

*Individual behavior: cooperation conditional on faith*

More generally, whether or not a player cooperates depends on the expectations about the behavior of the other player. As a measure of trust,  $q$  denotes the subjective probability a player attaches to the contingency that the other player cooperates. A risk-neutral player would act cooperatively if the expected pay-off of cooperation exceeds the expected pay-off from defection:  $q2(B - C) + (1 - q)(-C) > q(B - C)$ . We can rewrite this inequality as<sup>17</sup>

$$q > \frac{C}{B} = \frac{1}{1 + R} \quad (2)$$

This inequality shows that cooperative behavior is conditional on faith  $q$  (i.e., having enough trust that the other player cares for your interests).<sup>18</sup> In particular, people cooperate if the actual risk premium for credit risk  $R \equiv \frac{B-C}{C}$ , which can be interpreted as hope, exceeds the required risk premium on account of a lack of trust  $\frac{(1-q)}{q}$ .<sup>19</sup> Both hope and trust thus lead to cooperation. Indeed, hope raises the actual risk premium, whereas trust reduces the required risk premium for investing in cooperation.<sup>20</sup>

---

<sup>17</sup> We talk about negative intrinsic reciprocity if the decisionmaker attaches a negative weight to the welfare of agents who hurt them. The decisionmaker than in fact hates these agents rather than being indifferent towards them. In the presence of both positive and negative intrinsic reciprocity (both with a weight of unity), the inequality that determines whether an agent cooperates becomes  $q2(B - C) + (1 - q)(-C - B) > q(B - C)$ , which can be rewritten as  $q > \frac{1}{2} \left(1 + \frac{1}{1+R}\right)$ . Hence, the minimum level of trust required to induce the agent to cooperate exceeds  $\frac{1}{2}$  and exceeds the trust level required for cooperation with only positive reciprocity.

<sup>18</sup> Strong reciprocity may also operate on the group level. An individual who is strongly reciprocal will contribute to the public good if the following inequality is met  $\frac{B}{N}(1 + M - (N - M - 1)) > C$ , where  $M+1$  is the number of cooperating individuals (including the strongly reciprocal person whose choice we are considering). We can rewrite this inequality as  $q > \frac{1}{1+R}$ , where  $q \equiv \frac{M+1}{N}$  is the proportion of the population that is cooperating. This inequality for the public-goods game has the same form as inequality (2) for the prisoner's dilemma—with one important difference: in the public-goods game  $q$  denotes the proportion of the population that cooperates, whereas  $q$  in the prisoner's dilemma represents the probability that the other player cooperates.

<sup>19</sup> The required risk premium would be higher if the decisionmaker would be risk-averse or loss-averse. The reason is that a risk-averse individual would attach a lower subjective (welfare) weight to the upside of cooperation than the objective probability  $q$  and thus a higher subjective (welfare) weight to the downside of cooperation.

<sup>20</sup> If people attach a weight  $0 < p < 1$  (rather than 1) to the material pay-offs of the people who act cooperatively toward them, expression (2) becomes  $pq > \frac{1}{1+R}$ . This inequality shows that love ( $p$ ) and trust ( $q$ ) are closely related if people are strongly reciprocal: both the love parameter ( $p$ ) and the trust parameter ( $q$ ) have similar positive effects on cooperation. Indeed, love ( $p$ ), faith ( $q$ ) and hope ( $R$ ) all contribute to cooperation.

*Group behavior**The dynamics of trust on a group level and over time*

On an individual level and in the short run, the level of trust is exogenously given and the causality runs from trust (i.e. the immaterial world) to works (i.e. the material world); see the arrow from the top to the bottom at the right-hand side of figure 3. On a group level and in the longer run, however, trust can be considered an endogenous collective good that responds to actual behavior of others in the group. Indeed, trust is built on our actual experiences in a household (i.e. a group): the causality thus runs from actions in the observable material sphere to the spiritual sphere (i.e. subjective beliefs and values); see the arrow from the bottom to the top at the left-hand side of figure 3.

**Figure 3. Trust and cooperation.**

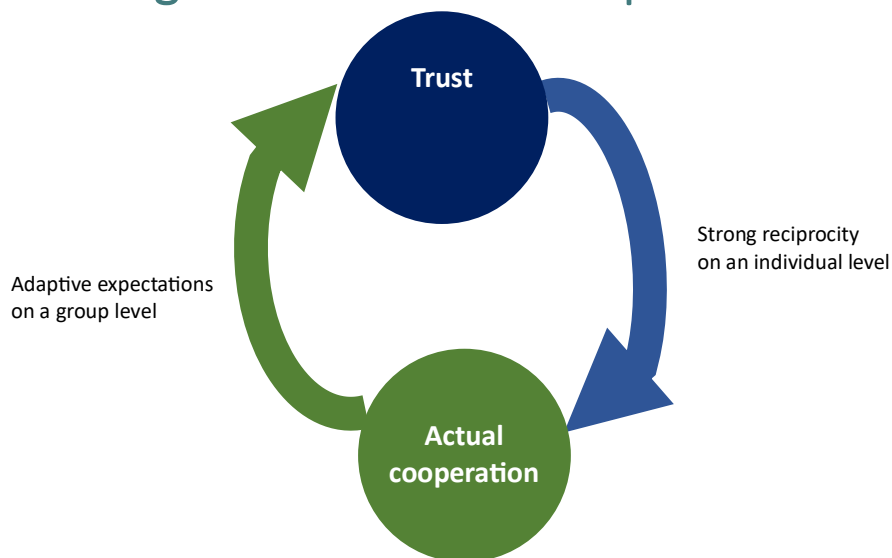
*Positive feedback between trust and cooperation*

Figure 3 shows that trust and cooperation are mutually dependent in a group of people with strongly reciprocal preferences. On the one hand, people cooperate and value the interests of others if they can trust others to protect their interests (see the arrow from the top to the bottom at the right-hand side of figure 3). On the other hand, trust is built up and maintained if people are indeed seen to be trustworthy by actually cooperating in practice (see the arrow from the bottom to the top at the left-hand side of figure 3). Reciprocity thus characterizes the relationship between trust and cooperation.

*Positive and negative spirals*

The positive feedback between trust and cooperation can cause trust, mutual regard and mutual service to grow so that an economy becomes more productive over time. However, the

feedback also operates in the reverse direction: fear and selfishness strengthen each other so that the productivity of an economy goes from bad to worse. Accordingly, the feedback between trust, appreciation and cooperation can cause both positive and negative spirals and render both trust and fear self-fulfilling.

#### *Two Nash equilibria*

The positive feedback between trust and cooperation leads to two pure strategy Nash-equilibria<sup>21</sup> in a prisoner's dilemma (or public goods game) with players exhibiting strong reciprocity as a social preference (see table 3 and figure 3). In a so-called Nash equilibrium, expectations coincide with actual behavior. In the cooperative equilibrium all players cooperate because they trust and value each other's material pay-offs. Trust in cooperation is then confirmed in actual behavior. But the situation in which nobody cooperates is a Nash equilibrium too. In this so-called non-cooperative equilibrium, nobody internalizes externalities because people do not trust each other. The fear that others are not looking after their interests is confirmed in the actual non-cooperative behavior of people.

#### *Normative implications: Nash equilibria are Pareto ranked*

For both players, the cooperative equilibrium yields a higher pay-off than the non-cooperative equilibrium. A game such as in Table 3 (in which there are two Nash equilibria that can be Pareto ranked) is called a coordination game of the stag hunt type. The existence of two equilibria is consistent with our experience: humans are able not only to cooperate but also to hurt each other. Even though we all would like to cooperate, cooperation often does not happen. Cooperation is valuable but difficult.

#### *Adaptive expectations about the cooperation of others*

Adaptive expectations about cooperation in the group can be modelled as follows (here the subscript is a time subscript indicating the period):<sup>22</sup>

$$q_{t+1} = q_t + \alpha(w_t - q_t). \quad (3)$$

Here,  $q_t$  stands for the expectation at time  $t$  about whether others in a group will cooperate in the future.  $w_t$  represents the proportion of the population that actually cooperates (so that a proportion  $1 - w_t$  defects) and is thus a measure of cooperation within the group. The parameter  $\alpha$  indicates how rapidly subjective expectations respond to actual behavior. This equation shows that expectations about cooperation are adapted upwards (downwards) if more (fewer) people cooperate than expected.

---

<sup>21</sup> Pure strategies are strategies that are not stochastic. Allowing for so-called mixed strategies in which players cooperate with a certain probability  $0 < p < 1$  makes even more equilibria possible. In these mixed strategy equilibria, the level of social capital  $p$  can be anywhere between zero and 1.

<sup>22</sup> A literature exists on the dynamics of mutual expectations of players in sequential games. See, for example, Dufwenberg and Kirchsteiger (2004).

*Dynamics: bifurcation point*

If the development of trust is described by equation (3), which of the two equilibria emerges depends on the initial level of social capital  $q_0$ . In particular, the trust level  $\frac{1}{1+R}$  at the right-hand side of equation (2) is the bifurcation point. In particular, if the initial trust level exceeds this threshold (i.e.  $q_0 > \frac{1}{1+R}$ ), people cooperate (i.e.  $w_0 = 1$ ) and trust thus grows. If the initial trust level is below the bifurcation point for trust  $\frac{1}{1+R}$ , in contrast, people do not cooperate (i.e.  $w_0 = 0$ ) and trust therefore declines.<sup>23</sup>

*The initial team spirit of social capital as a gift of grace*

The stag-hunt game shows that mutual trust is the most important capital of a group. This collective good of the team spirit (i.e. the spirit that governs the group) is the social capital of the group involving mutual trust, mutual love, and mutual service within a group. The collective spirits (or team spirits) involving shared beliefs and common values can also be called culture or common consciousness.<sup>24</sup> Indeed, these shared beliefs and values involve the immaterial world of subjective beliefs and subjective values. To be born into a family or a country with a culture that has a high level of social capital can be viewed as a gift of grace. Indeed, a larger stock of social capital makes it easier to protect and supply public goods, such as the natural environment. Differences across countries in levels of social capital in fact explain inequality between countries (Knack and Keefer, 1997, and Xue, Reed and van Aert, 2024).

*Cooperation as an unstable equilibrium*

The stag hunt game shows that human cooperation is possible but inherently unstable. Trust can take years to build but can evaporate quickly.<sup>25</sup> By eroding social capital, a few free-riders can thrust a whole household into a so-called trust fall. The cooperative equilibrium is then turned into a non-cooperative equilibrium because the contagious virus of fear and selfishness goes viral.<sup>26</sup> Compared to the direct damage of non-cooperative behavior, the indirect effect of that defection on the collective good of social capital may in fact be much more damaging in terms of lost material value. Indeed, if social capital declines below the bifurcation value, a household becomes entrapped in a non-cooperative equilibrium in which fear and selfishness are mutually reinforcing.

*Humans as relational beings: Herd behavior and amplification of shocks*

<sup>23</sup> This illustrates how the so-called Matthew effect applies to trust in a group (Matthew 25:29): "For to everyone who has will more be given, and he will have abundance; but from him who has not, even what he has will be taken away."

<sup>24</sup> Harrari (2017) calls these common beliefs and values intersubjective realities.

<sup>25</sup> Loss aversion is one reason why trust evaporates more quickly than it can be built up. In that case, the adjustment coefficient  $\alpha$  in expression (3) is larger for negative than positive values of  $(w_t - q_t)$ .

<sup>26</sup> Equation (2) shows that for a cooperative equilibrium in a public goods game the share of free-riders should not exceed  $1-q = \left(\frac{R}{1+R}\right)$ .

The case of a trust fall as a consequence of a few free-riders illustrates how social preferences of the strong reciprocity type lead to more strategic interactions and external effects among people than self-interested preferences do. Humans, in effect, become more relational beings if they demonstrate preferences of strong reciprocity. They affect each other's beliefs, preferences, welfare, and behavior. Indeed, preferences are social (or relational) not only because people value external effects on others but also because their preferences are shaped by the actions of others, namely through the effect of these actions on the collective good of social capital.

#### *Keynesian economics*

The instability of group behavior due to contagious emotions is reminiscent of the macroeconomic thinking of Keynes; see Camerer and Fehr (2006) and Akerlof and Shiller (2009). Keynes stressed the importance of narratives and emotions that affect the confidence of consumers and investors. These contagious emotions, or "animal spirits," can shift entire economies into negative spirals of fear or positive spirals of euphoria. Temporary shocks can thus trap economies permanently and result in positive feedbacks amplifying shocks.

## **5. Reflections and conclusions**

This paper has developed a more relational anthropology than the selfish homo oeconomicus. This anthropology involves the concept of strongly reciprocal preferences developed in evolutionary biology, experimental economics, and evolutionary psychology. In particular, people derive relational goods from giving material goods to people with whom they identify themselves. By serving others, they enjoy both material and relational goods.

Understanding the human being as a homo amans who likes to reciprocate love can help economists to more accurately explain human motivation, rational behavior and well-being. It can also help to explain the potential contribution of religion to well-being and cooperation. In addition, understanding the two Nash equilibria in the stag hunt game can help theologians to distinguish between the kingdom of God (or of light) and the kingdom of this world (or of darkness). In the kingdom of God, trust in a good God and love for neighbors prevail, so that cooperation yields material and social goods for all. In the absence of trust, however, people live in the kingdom of darkness, where poverty and indifference reign. In that event, humans miss their calling to serve one another and flourish. Indeed, the connection between trust in a good, loving God, on the one hand, and cooperation as mutual service, on the other, seems a fruitful avenue for future research.

Models with strongly reciprocal preferences explain why human cooperation is fragile and why sin (defined as lack of trust and love) is contagious. The case of a trust fall due to a few free-riders eroding the cooperative team spirit of trust and love illustrates how social preferences of the strong reciprocity type give rise to strategic interactions and external

effects. In fact, this is reminiscent of Keynesian economics, in which contagion can amplify temporary shocks in beliefs as “animal spirits.”

## References

- Akerlof, George, and Robert Shiller. 2009. *Animal Spirits: How Human Psychology Drives the Economy and Why it Matters for Global Capitalism*. Princeton: Princeton University Press.
- Andreoni, James. 1990. “Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving.” *The Economic Journal* 100: 464–77.
- Bovenberg, Lans, and Bas van Os. 2025. “Hope and Love in Sequential Games.” *Journal of Economics, Theology and Religion* 5, no. 1: 81–95.
- Bowles, Samuel. 2004. *Microeconomics, Behavior, Institutions and Evolution*. Princeton: Princeton University Press.
- Bowles, Samuel. 2016. *The Moral Economy: Why Good Incentives Are No Substitute for Good Citizens*. New Haven: Yale University Press.
- Bruni, Luigino, and Luca Stanca. 2008. “Watching Alone. Happiness, Relational Goods and Television.” *Journal of Economic Behavior and Organization* 65: 506–28.
- Camerer, Colin F., and Ernst Fehr. 2006. “When Does ‘Economic Man’ Dominate Social Behavior?” *Science* 311: 47–52.
- Dhami, Sanjit. 2019. *The Foundations of Behavioural Economic Analysis: Other-Regarding Preferences*. Oxford: Oxford University Press.
- Dufwenberg, Martin, and Georg Kirchsteiger. 2004. “A Theory of Sequential Reciprocity.” *Games and Economic Behavior* 47: 268–98.
- Fehr, Ernst, and Urs Fischbacher. 2003. “The Nature of Human Altruism.” *Nature* 425: 785–91.
- Gintis, Herbert. 2000. “Strong Reciprocity and Human Sociality.” *Journal of Theoretical Biology* 206: 169–79.
- Harari, Yuval Noah. 2017. *Sapiens: A Brief History of Mankind*. London: Vintage UK.
- Knack, Stephan, and Philip Keefer. 1997. “Does Social Capital Have an Economic Payoff? A Cross-Country Investigation.” *Quarterly Journal of Economics* 112: 1251–88.
- Kosfeld, Michael. 2020. “The Role of Leaders in Inducing and Maintaining Cooperation: The Conditional Cooperative Strategy.” *The Leadership Quarterly* 31, no. 1.
- Lewis, C.S. 1960. *The Four Loves*. London: Geoffrey Bles.
- Medema, Steven G. 2020. “The Coase Theorem at Sixty,” *Journal of Economic Literature* 58, no. 4: 1045–128.
- Nullens, Patrick, and Jermo van Nes. 2022. “Towards a Relational Anthropology Fostering an Economics of Human Flourishing.” In *Relational Anthropology for Contemporary*

- Economics*, edited by Jermo van Nes, Patrick Nullens and Steven van den Heuvel, 9–30. Cham: Springer.
- Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83: 1281–302.
- Seabright, Paul. 2010. *The Company of Strangers: A Natural History of Economic Life*. Princeton: Princeton University Press.
- Smerilli, Alesandra. 2007. "'We-Rationality:' A Non-Individualistic Theory of Cooperation." *Economia Politica* 3: 407–26.
- Sugden, Robert. 2018. *The Community of Advantage: A Behavioral Economist's Defense of the Market*. Oxford: Oxford University Press.
- Xue, Xindong, W. Robert Reed, and Robbie C. M. van Aert. 2024. "Social Capital and Economic Growth: A Meta-Analysis." *Journal of Economic Surveys*: 1–38.