

# How Can I Trust You? Encounters with Carl Rogers and Game Theory

Len Fisher

## 1 Introduction

In Smith's novel *Daughters of Darkness* (1996), her ruthless character Ash is asked "You always look after Number One, don't you?" "Doesn't everybody?" he replies with surprise.

Although Ash doesn't know it, he is following in a long philosophical tradition that began with Socrates two and a half thousand years ago. Socrates took it as an axiom that "everyone seeks what is most serviceable to oneself or what is in one's own self-interest".<sup>1</sup> Prominent followers in that tradition have included Niccolò Machiavelli (1532), who argued that a wise ruler ought never to keep faith when by doing so it would be against his interests, and the philosopher Thomas Hobbes (1651), who saw rational self-interest as the cardinal human motive.

Nearly four centuries later, the pursuit of rational self-interest remains a fundamental cornerstone of many Western societies. It forms the foundation of classical free market economics, for example, in which economists view us as members of the disinterestedly rational species *Homo economicus*, making informed decisions based solely on the best outcomes for ourselves (Anon 2005). Some philosophers have even attempted to justify it on moral grounds. Ayn Rand (1964), for example, sees rational self-interest as the "proper moral purpose of a person's life".

Psychologists and sociologists are well aware that this simple picture fails to acknowledge the complexity of our motivations and under-emphasises the role of such factors as public spiritedness, empathy, commitment and justice (Miller

---

<sup>1</sup> Socrates' basic axiom appears in several different guises in the Socratic dialogues reported by Plato. This particularly clear paraphrase is that which is used on the Lander University philosophy course website (<http://philosophy.lander.edu/ethics/socrates.html>) under the heading "Socratic Paradox".

---

L. Fisher (✉)  
School of Physics, Tyndall Avenue, Bristol BS8 1TL, UK  
e-mail: len.fisher@bristol.ac.uk

1999). There is no doubt, though, that the philosophy of looking after Number One runs deeply through Western culture.

The personal and social consequences of looking after Number One are the subject of *game theory*, which looks at our social interactions as games that we play<sup>2</sup> for the highest personal rewards with the lowest possible risk. Game theory was developed in the 1940s, first by the brilliant Princeton mathematician John von Neumann (von Neumann and Morgenstern 1944), thought by some people to have been the model for the remote and impersonal Dr. Strangelove, and later by the mathematical genius John Nash (1951), known to many as the schizophrenic anti-hero of the film “A Beautiful Mind”.

These mathematical antecedents are no accident, because game theory is concerned with calculating the odds of success or failure for the various social strategies that we might adopt. Until John Nash came along, however, no one had any idea that game theory would expose a social paradox that has been staring us in the face since the dawn of human society—a paradox which means that the use of “rational self-interest” as a guide to social interactions can often land us in situations where self-interest is the last thing that is being served.

The paradox concerns situations where cooperation would serve everybody’s interests, but where the logic of self-interest dictates that an individual can do even better by putting his or her own interests above those of the group and breaking the cooperation (in game theory parlance, *defecting*). What is sauce for the goose is sauce for the gander, however, and when all of the individuals in the group use the same ironclad logic, cooperation collapses and everyone ends up worse off than if they had maintained the cooperation in the first place.

This vicious logical paradox can affect us in many real-life situations, from divorce to war, from the breakdown of individual relationships to global problems such as pollution, resource depletion and climate change; so many, in fact, that it has been proposed as the basic problem of society, since our efforts to live together in a cooperative and harmonious way are so often undermined by it.

But is it a necessary paradox? Or does it arise because we live in an increasingly depersonalised society, where it has become essential to look after Number One, or risk going under? Is there some way that we can change our approach to social interactions and so avoid the paradox?

It has turned out that the key to resolving the problems exposed by game theory is the evocation and development of *trust* (Fisher 2008). Game theorists have so far tackled this question by developing strategies based on the logic of self-interest. Enter Carl Rogers. Rogers’ pioneering work on the spontaneous evolution of trust and acceptance in encounter groups (Rogers 1970) and on the evocation of trust through unconditional positive regard in the person-centred approach (Rogers 1942, 1951; Wilkins 2010), offers possibilities for a very different approach.

---

<sup>2</sup> Note that these are *not* the psychological games that Eric Berne refers to in *Games People Play* (1964). Berne does not refer to mathematical game theory in his book, even though it was written 20 years after formal game theory had been established.

My aim in this chapter is to help open a dialogue between psychologists and psychotherapists on the one hand (focussing particularly on the person-centred approach) and game theorists on the other hand<sup>3</sup> to investigate further development of trust and cooperation in human relationships, and how it can help to resolve the paradoxes exposed by game theory. These two groups approach the problem from very different viewpoints, using different axioms and modes of thinking. I believe that it is vitally important for each to understand the other, and to be informed by the other, if we are to make real progress.

My approach is frankly speculative and designed to stimulate discussion rather than offer dogmatic solutions to problems that have been with us since the dawn of civilisation. If I succeed in getting experts from either side of the fence to think about the possibility of a two-pronged attack on these problems, or at least the possibility that there is more than one way of looking at them, then I will have done my job.

## 2 The Prisoner's Dilemma

I begin by looking at the problem of cooperation from the game theorist's point of view. That point of view is encapsulated in the now-famous parable of *The Prisoner's Dilemma*, which the Princeton mathematician Albert Tucker invented when he was asked to explain game theory to a group of psychologists at Stanford University. As recounted later by his colleague Kuhn (1994):

Tucker was on leave at Stanford in the Spring of 1950 and, because of the shortage of offices, he was housed in the Psychology Department. One day a psychologist knocked on his door and asked what he was doing. Tucker replied: "I'm working on game theory", and the psychologist asked if he would give a seminar on his work. For that seminar, Al Tucker invented the Prisoner's Dilemma (p. 161).

The story has since appeared in various incarnations. In one of them, two thieves (let's call them Bernard and Frank, after two of the conspirators in the Watergate scandal) have been caught by the police, but the prosecutor has only enough evidence to put them behind bars for 2 years on a charge of carrying a concealed weapon, rather than the maximum penalty of 10 years that they would get for burglary. So long as they both plead "not guilty", they will both get only 2 years, but the prosecutor has a persuasive argument to get them to change their pleas.

---

<sup>3</sup> I am not a practitioner in either field, but a scientist who has written and lectured extensively on game theory in everyday life in an effort to make its socially important discoveries more widely known (see, for example, my *Rock, Paper, Scissors: Game Theory in Everyday Life* (Fisher 2008)). My appreciation of the person-centred approach has come principally from my wife Wendella, a counsellor who uses and has great experience with the person-centred approach, and with whom I have attended a number of encounter groups.

He first approaches Bernard in his cell and points out that if Frank pleads guilty but Bernard doesn't, Frank will receive a reduced sentence of 4 years for pleading guilty, but Bernard will get the maximum 10 years. So, Bernard's best bet, if he believes that Frank will plead guilty, is to plead guilty as well, so as to receive 4 years rather than ten. "Furthermore" says the prosecutor "I can offer you a deal that if you plead guilty and Frank doesn't, you can go free for turning state's evidence!"

No matter what Frank does, it seems that Bernard will always do better for himself by pleading guilty. The logic seems irrefutable—and it is. The trouble is that the prosecutor has made the same offer to Frank, who has come to the same conclusion. So, they both plead guilty—and they both end up in jail for 4 years, rather than the 2 years that they would have received if they had both kept their mouths shut.<sup>4</sup>

Tucker's story of the Prisoner's Dilemma goes straight to the heart of the paradoxes that can arise when we use rational self-interest as our guide to action. It has struck resonances with many people, and literally thousands of articles and dozens of books have been devoted to examining the consequences of its insidious logic and to proposing solutions to the paradox. Not all of these books and articles have been by game theorists. As the journalist William Poundstone recounts in his 1993 book *Prisoner's Dilemma*, philosophers, religious leaders, politicians, psychologists, sociologists, and the inevitable collection of cranks have all had their say.

### 3 Trust

One of the major conclusions from all serious contributors to the debate is that the key to a solution lies in *trust*. If each of the prisoners in Tucker's story had been able to trust the other not to give them away, their problem would have been solved. But how can such trust be achieved?

Promises are clearly insufficient, as the *dramatis personae* in Puccini's opera *Tosca* discover to their cost.<sup>5</sup> The heroine (Tosca) is faced with an unenviable choice. Her lover (Cavaradossi) has been condemned to death by the corrupt police chief Scarpia. Tosca is left alone with Scarpia, who thinks that he is on to a good thing when he offers to have the firing squad use blank bullets if Tosca will let him have his wicked way with her. Tosca agrees—but is her commitment credible?

Scarpia thinks it is, because he has Tosca on her own in a room from which there seems to be no escape. But Tosca has spied a knife on the table and has

---

<sup>4</sup> If you think that this little story has uncomfortable parallels with the U.S. practice of plea bargaining, you are dead right. This is why plea bargaining has been made illegal in many countries.

<sup>5</sup> This example is due to Bill Poundstone (1993).

worked out that she can win both ways by agreeing to Scarpia's proposal, but actually stabbing him when he comes close.

Unfortunately for Tosca, Scarpia's commitment wasn't credible either! It was no more than an empty verbal contract and (as Hollywood producer Sam Goldwyn is supposed to have said), verbal contracts aren't worth the paper they are written on. In fact, Scarpia has worked out that *he* can win both ways by having his way with Tosca, but not really telling the firing squad to use blank bullets. The upshot is operatic mayhem. Cavaradossi dies, Scarpia dies, and when Tosca finds out what has happened, she flings herself off a castle parapet and she dies too. Everyone is a loser, as is often the way with opera.

Everyone is a loser in real life as well when promises can't be trusted, whether the promise has come from a partner, a politician or a passerby. But how are we to achieve such trust?

The game theorist's answer lies in *credible commitment* to the promise that has been given. Such commitment can be evoked by using the logic of self-interest if (a) the person offering the commitment puts themselves in a position where it is obvious to the other party or parties that the commitment is irreversible, or (b) it can be seen by the other party or parties that it would be too costly for the person offering the commitment to change their mind later.

How might do such strategies work in practice? Below I offer a series of examples. They show that the logic of self-interest can quite often produce practical strategies for demonstrating credible commitment, but that all too often there is a loophole which allows one or other party to cheat on the cooperation for personal advantage and without having to pay too heavy a penalty.

### ***3.1 Deliberately Cutting Off Your Escape Routes***

There are three broad ways to do this, each scarier than the last.

#### **3.1.1 Use a Mandated Negotiating Agent**

With a legally binding contract, that agent is the law. But there are many "contracts" that we enter into which are not legal contracts, but which are contracts nonetheless. When my brother and I divided up the household jobs between us, our verbal agreement was a contract, and it was enforced because we had a mandated negotiating agent—our father!

If the two prisoners in Tucker's story had each had friends on the outside who could be relied upon to punish the other for giving him away, these would also be acting as "mandated negotiating agents" and would have saved the day.

The loophole with contracts is that they can often be renegotiated (witness what happens when countries or business firms declare bankruptcy) and can in any case be difficult and costly to enforce.

### **3.1.2 Cut Off Communication**

We've all done it. We do it whenever we post a letter, press the "send" button for an e-mail, turn off our mobile phones, or even when we have written our wills. Once we've done it, that's it. We've made a commitment, and that commitment is credible because there is no going back. But we can apologize for the e-mail, change our wills up to the moment that we die and say that we "forgot" to turn our phone on. There's usually a way out.

### **3.1.3 Burn Your Bridges**

Cutting off communication is one way to "burn your bridges", but there are many others. A striking example is that of the Spaniard Hernando Cortés, who led an expeditionary force to invade Mexico in 1519. Cortés scuttled his ships in full view of the on-looking Aztecs, thus making it impossible for his force to retreat, and demonstrating to the Aztecs his commitment to remain.

Two friends of mine found another way to burn their bridges when they decided to do a parachute jump. Both of them got an attack of nerves, with each saying to the other "if you go first, I'll follow you". Neither would really trust the other to follow until they hit on the idea of offering credible commitment by each taking a grip on the other's wrist, so that when one jumped, the other was forced to follow.

Of all the logical strategies for credible commitment, burning your bridges is the one with the most force.

## ***3.2 Making It Too Costly for You to Change Your Mind Later***

There are many possible strategies. Here are four major ones:

### **3.2.1 Put Yourself in a Position Where Your Reputation Will Be Damaged if You do not Deliver**

This can be a powerful strategy in personal relationships, because letting down others in the group can do you future damage when they then fail to trust or accept you. It seems to be much less powerful in politics, where promises made in order to gain power are frequently broken later.

A particularly important, and much studied, possibility is to use *repeated interactions*. If you know that you are going to have to cooperate with someone again in the future, you are much less likely to cheat on a promise or renege on a bargain. But people still do.

### 3.2.2 Move in Steps

Breaking a promise or threat into a series of steps means that, when you get towards the end, most of the promise or threat will have been fulfilled, as happens when homeowners or developers pay builders at the end of each completed phase of a project. But there is a trap here. If you *know* that it is the last step, you may be tempted to renege. A developer, with the project completed, may refuse the last payment, leaving the builder out of pocket, or with the stress and cost of taking the developer to court. A tenant may skip without paying the last month's rent, as has happened to me as a landlord more than once. The message is clear; make the steps (or at least, the last few steps) as small as possible so as to minimise the risk of loss. In the last month of a lease, for example, make the payments weekly rather than monthly, or ask for payment in advance.

### 3.2.3 Enter Into a Contract

Some contracts are binding, as Faust discovered when he entered into a contract with the devil. But most contracts are not so binding and can be subject to renegotiation. To make them stick, it often needs something extra, such as a penalty clause. The person or body who enforces the clause must also have a good reason to stick to their responsibility. Penalty clauses are of little use if a local planning officer can be bribed into passing a shoddy piece of building work, for example, even though that work does not meet the standards of the contract.

### 3.2.4 Use Brinkmanship

"I'll shoot" screams a man standing at the counter of a bank "unless you pass over that bag full of money!" How realistic is his threat? It doesn't really matter, because the outcome will be so drastic if he carries it out. That's the essence of brinkmanship, a term coined by U.S. Presidential candidate Adlai Stevenson at the height of the cold war in 1956. Stevenson used it to criticise Secretary of State John Foster Dulles for "bringing us to the edge of the nuclear abyss". It demonstrates credible commitment by making the cost of escape too high, although it is certainly the least likely of the lot to lead to genuine cooperation!

## 4 Enter Carl Rogers

Logic is not the only way to generate credible commitment. Close involvement within a group can do just as good a job, even between strangers, as Carl Rogers discovered when he studied the behaviour and evolution of encounter groups in the

1960s. Here are some of his observations that I believe are relevant to our discussion here (from Rogers 1970; pages 18, 14, 16, 40, 28 and 50, respectively):

... the soil out of which this demand [for encounter groups] grows has two elements. The first is the dehumanization of our culture ...

A climate of mutual trust develops [in encounter groups] out of [the] mutual freedom to express real feelings, positive and negative.

... one of the most common developments is that a sense of trust slowly begins to build ...

One member ... speaks of the “commitment to relationship which often developed on the part of two individuals ...”

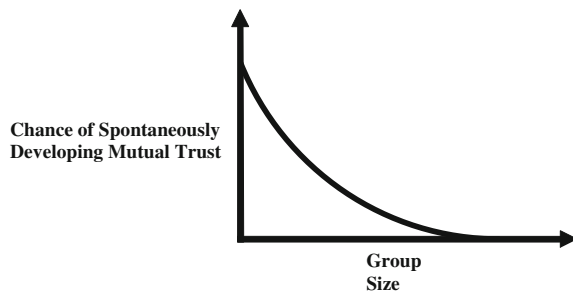
One of the most fascinating aspects of any intensive group experience is ... the manner in which a number of the group members show a natural and spontaneous capacity for dealing in a helpful, facilitating and therapeutic fashion with the pain and suffering of others

... the group seems like an organism, having a sense of its own direction even though it could not define that direction intellectually.

In other words, many people feel alienated and isolated in our wider culture and compelled in self-defence to adopt the dehumanizing “rational self-interest” approach to handling their interactions with most other people—an approach that can lead to the serious problems exposed by game theory. When people are given an opportunity to get together in an initially unstructured group, however, their human qualities come to the fore, trust and mutual support emerge, and the group eventually takes on a dynamic of its own.

This is obviously a very broad generalisation (albeit one that is supported by research), and I put it forward here as a catalyst for discussion rather than as a dogmatic assertion. It shows, at least, that there are at least two possible routes to credible commitment—via strategies based on the logic of self-interest, or through spontaneous group dynamics. In the latter case, I would suggest that the potential for the development of the mutual trust necessary to overcome the paradoxes of game theory depends very much on the size of the group, as illustrated schematically in Fig. 1:

**Fig. 1** Potential for mutual trust development sufficient to overcome the paradoxes of game theory as a function of group size



The question is: “Could either of these routes to credible commitment (used separately or in tandem) help us in practice to avoid or escape from situations such as that exemplified by the Prisoner’s Dilemma?”



## 5 The Seven Deadly Dilemmas

Game theorists have identified seven basic situations [which I call *The Seven Deadly Dilemmas* (Fisher 2008)],<sup>6</sup> where the use of rational self-interest takes us to a less-than-ideal place. In addition to The Prisoner's Dilemma, there are

- *The Tragedy of the Commons*, where individuals who share a common resource are each tempted to take more than their fair share. When they all follow this strategy, however, the resource becomes overused and can even disappear, as witness the collapse of many fisheries.
- *The Free Rider* problem (a variant of the Tragedy of the Commons), which arises when some people in a community take advantage of a group resource without paying for it.
- *Chicken* (also known as *Brinkmanship*), where each side tries to push the other as close to the edge as they can, with each hoping that the other will back down first. It can arise in situations ranging from someone trying to push into a line of traffic to confrontations between nations that could lead to war, and which sometimes do.
- *The Volunteer's Dilemma*, in which someone must make a sacrifice on behalf of the group, but if no one does, everyone loses out. Each person is hoping that someone else will be the one to make the sacrifice, which could be as trivial as making the effort to put the garbage out, or as dramatic as one person sacrificing his or her life to save others.
- *The Battle of the Sexes*, where two people have different preferences, such as a husband who wants to go to a ball game while his wife would prefer to go to a movie. The catch is that each would rather share the other's company than pursue their own preference alone. But how can they make the decision?
- *Stag Hunt*, a situation where cooperation between members of a group gives them a good chance of success in a risky, high-return venture, but where an individual can win a guaranteed, but lower, reward by breaking the cooperation and "going it alone".

Here, I examine each of these dilemmas in turn and ask whether logic-based and/or person-centred routes to credible commitment could help to resolve the dilemma.

### 5.1 The Prisoner's Dilemma Revisited

The prisoners in Tucker's story may have been able to escape their dilemma if each of them had had friends on the outside, willing and able to punish the other

---

<sup>6</sup> Formal logical analysis requires a concept known as the *Nash equilibrium*, which is not difficult to master, but which I have avoided using in the context of this chapter. A description of the seven deadly dilemmas in terms of the Nash equilibrium may be found in my book *Rock, Paper, Scissors* (Fisher 2008).

for giving their friend away. If each prisoner regarded the threat as credible, then rational self-interest would dictate that each should keep his mouth firmly shut. As with so many instances of rational self-interest, however, there could be a loophole. The ratting prisoner's friends might protect him, for example, or the authorities might provide protection, or even a new identity.

Practical experience has also shown that criminals are unlikely to rat on each other if they are members of the same criminal gang. This could be due to the development of trust within the group, the fear of loss of reputation within the group, the anticipation of repeated interactions, or a combination of all of these.

But this story is not just about criminals and criminal gangs. The same principles apply to *any* situation where one individual might be tempted to cheat on cooperation with another for personal gain, whether it be within a marriage, a community or a business arrangement.

### 5.1.1 Credible Commitment and the Prisoner's Dilemma

"Credible commitment" via a marriage settlement, a contract or fear of loss of reputation provides one possibility for evoking trust, although experience has shown that these logic-based mechanisms are not always reliable. Mutual membership of a group where trust has evolved spontaneously provides a more reliable mechanism—witness the degree of support that members of small communities often offer each other, not to mention church groups or organisations such as the Masons (which provided great support to my mother after my father died).

There is also another possibility—mutual respect based on Rogers' principle of unconditional positive regard, where the reward works both ways. "I have found it highly rewarding", Rogers wrote in his 1961 essay *This Is Me* "when I can accept another person". The person who has received the unconditional acceptance is unlikely to cheat on the person who has offered it, if for no other reason than the risk that the source of acceptance might be cut off. The person who has offered the unconditional acceptance is also unlikely to cheat, not just because of the reward that Rogers wrote about, but simply because he or she *has* offered unconditional acceptance.

Unfortunately, these considerations carry less weight with my next Dilemma—The Tragedy of the Commons, which game theorists have proved may be formally viewed as a set of Prisoner's Dilemmas played out by all of the different pairs of individuals within a group.

## 5.2 *The Tragedy of the Commons*

This scenario was brought to public attention by the Californian ecologist and game theorist Garrett Hardin in a 1968 essay with the above title, although philosophers have been worrying about it since the time of Aristotle. Hardin

illustrated it with the parable of a group of herders each grazing their own animals on common land, with one herder thinking about adding an extra animal to his herd. An extra animal will yield a tidy profit, and the overall grazing capacity of the land will only be slightly diminished, so it seems perfectly logical for the herder to add an extra animal. The tragedy comes when all the other herders think the same way. They all add extra animals, the land becomes overgrazed and soon there is no pasture left.

The intractable paradox exemplified by the tragedy of the commons underlies family disagreements about inheritance, divorce settlements where the lawyers end up with the bulk of the proceeds, and choices about who should take responsibility for aged parents. On a wider scale, it is responsible for resource depletion, global warming and a host of other global problems (up to and including war).

At its heart lies an insidious logic. When just two people are involved, a gain for one is going to be an obvious loss for the other, and a balance may be struck. When many people are involved, however, the gain for an individual is palpably obvious, but the loss is spread across the group and can be so diluted as to become almost invisible.

We even see it with teaspoons. When a group of Australian medical epidemiologists started wondering about the way in which teaspoons were disappearing from the communal area of their office, they had a lot of fun at first dreaming up unlikely explanations. One was that the spoons had escaped to a planet entirely populated by spoon life-forms, there to live an idyllic existence where they were not being dunked head-down in cups of hot tea or coffee. Another was *resistance*—the belief that inanimate objects have a natural antipathy towards humans and are forever trying to frustrate us, in this case by hiding when they are most wanted, in the manner of single socks in a washing machine.

The true explanation, of course, was that they were faced with a domestic version of the Tragedy of the Commons. “Teaspoon users”, they said “(consciously or otherwise) make decisions that their own utility [i.e. the benefit to themselves] is improved by removing a teaspoon for personal use, whereas everyone else’s utility is reduced by only a fraction per head (after all, there are plenty more spoons...). As more and more teaspoon users make the same decision, the teaspoon commons is eventually destroyed”.

It sounds funny when applied to teaspoons, but if you replace the word “teaspoon” by “land”, “oil”, “fish”, “forest”, or the name of any other common resource, you will soon see that some very serious global problems have their origins in this vicious circle of logic, which can make its unwelcome presence felt whenever profit goes to an individual person or group of people, but costs are shared by the community as a whole.

### 5.2.1 Credible Commitment and the Tragedy of the Commons

The Tragedy of the Commons is one of the most serious problems facing us in the world today. The failure of international agreements on fishing quotas, rainforest

preservation, global pollution and the like shows how difficult it can be to produce credible commitment by supposedly rational means.

There is no easy solution (in some cases, there may be no solution at all) in terms of the “rational” strategies for producing credible commitment as suggested by game theory. One possibility, however, that has been discussed in many contexts is that of “modularization” of the problem—in other words, breaking the system up into smaller, more self-sufficient units that are less dependent on each other. This has been suggested in the context of the international banking system (May et al. 2008), ecosystems (Allesina and Tang 2012), and social-ecological systems (Ostrom 2009). The key point is that the groups which make the decisions must be small enough to be able to perceive the costs to themselves of various strategies, rather than looking at the benefits and assigning to costs to some nebulous larger group—small communities where the members know and trust each other, villages rather than towns, local communities rather than central bureaucracies. It should work in theory, and Carl Rogers’ experience with groups suggests one of the reasons why it can work in practice [as it has been shown to do in many individual instances (Fisher 2008)]. Whether it can be made to work in face of the relentless trend towards social aggregation, and agglomeration is another matter.

### 5.3 *Free Rider*

The free rider problem applies to any situation where a resource that has to be paid for cannot easily be restricted to those who have paid for it. The problem can become especially acute when it comes to the care and use of communal resources. The Greek philosopher Aristotle was one of the first to point out its existence when he observed that “That which is common to the greatest number has the least care bestowed upon it”.

The Chinese authoress Aiping Mu provides a poignant modern example in her book “Vermilion Gate”, which tells the story of her growing up during the cultural revolution:

During the “storm of communization”, peasants put much less energy into working for the collective economy than for themselves, because the rewards were the same no matter how much or how little they worked, and no one could be bothered to take care of the collective property. The most painful experience was eating at the mass canteens, which were supposed to liberate women from daily cooking and hence to increase their productivity and increase the quality of life. The outcome was just the reverse.

Misled by the propaganda, peasants assumed that a life of abundance had begun, and they could eat their fill ... .. the peasants lost nearly everything, even their cooking utensils and food reserves ... .. When the famine ended ... one estimate put the number of deaths in rural China at 23 million (as cited in Fisher 2008, pp. 67–68).

### 5.3.1 Credible Commitment and the Free Rider

“Free riding” encompasses such actions as littering, fare-dodging, tax-dodging and illegal dumping on both small and large scales. As with the Tragedy of the Commons, it is difficult to deal with because the benefit goes to an individual, but the community shares the costs.

In this case, however, some of the logic-based strategies suggested by game theorists could work. For example, the threat of social disapproval for littering is so strong in Scandinavian countries that there is scarcely a problem, while the threat of punishment in Singapore brings a different sort of severe cost. “Free riding” is also much less prevalent in smaller communities, because repeated interactions with others bring an unacceptably high social cost.

Perhaps the Rogers approach could also help here in the form of a lesson: if each of us first learns to take responsibility for ourselves, and cherish ourselves as individuals, then we are more likely to take communal responsibilities and cherish others.

## 5.4 *Chicken*

“Chicken” is not just a game for teenagers. It is all around us, even at the very highest levels, as the philosopher Bertrand Russell pointed out during the Cold War between the U.S.S.R and the United States in the 1950s (Russell 1959):

Since the nuclear stalemate became apparent, the Governments of East and West have adopted the policy which Mr. Dulles calls “brinkmanship”. This is a policy adapted from a sport which, I am told, is practised by some youthful degenerates. This sport is called “Chicken!” . . . . As played by irresponsible boys, this game is considered decadent and immoral, though only the lives of the players are risked. But when the game is played by eminent statesmen, who risk not only their own lives but those of many hundreds of millions of human beings, it is thought on both sides that the statesmen on one side are displaying a high degree of wisdom and courage, and only the statesmen on the other side are reprehensible. This, of course, is absurd. Both are to blame for playing such an incredibly dangerous game. . . . (p. 30)

We are constantly playing games of chicken in our everyday lives, whether we are walking towards someone on a narrow sidewalk, hoping that someone else in a group will offer to buy the next round of drinks, or waiting for someone else to tell the boss that he’s got things wrong. Whoever makes the first move loses out, while the others gain, but if no one makes a move to resolve the situation, everyone loses out.

### 5.4.1 Credible Commitment and Chicken

How can we resolve such situations? “Credible commitment” to your threat is hardly an answer (especially if it involves commitment to nuclear warfare!), but there is a different approach—to coordinate your actions.

If both parties simultaneously move slightly aside when walking towards each other, for example, then perhaps neither of them need step down into the gutter. If people agree to talk to the boss as a group, no individual will be singled out. In 1961, for example, towards the end of the Cold War, Khrushchev and Kennedy defused the Cuban missile crisis by agreeing to make simultaneous moves—Khrushchev to remove the missiles, Kennedy simultaneously lifting the blockade (Fisher 2008).

Simultaneous moves need negotiation, which takes us straight back to trust and credible commitment—each side must trust the other to be committed to fulfilling their side of the bargain, and the commitment must be credible. Strategies based on the logic of self-interest to demonstrate credible commitment to the promise can and do work. But how much easier if the trust is already there through prior small-group interactions!

## 5.5 *The Volunteer's Dilemma*

The now-extinct Yagan Indians of Tierra del Fuego had a wonderful word for a situation that we have all experienced. The word was *mamihlapinatapai* and it means “looking at each other hoping that the other will offer to do something that both parties desire to have done but are unwilling to do themselves”. It was described in the 1993 Guinness Book of Records as “the most succinct word in any language”.

The Yagan Indians did not become extinct through *mamihlapinatapai*, but it certainly encapsulates what we now know as The Volunteer's Dilemma, where someone has to make a sacrifice on behalf of the group, but if no one does, then the whole group will suffer.

Being the volunteer, though, can require a courage amounting to heroism. When a grenade was lobbed into the middle of a platoon led by Sergeant Laszlo Rabel of the U.S. infantry, the platoon members would have died or been seriously injured if they had all stood back hoping that someone else would act. Sergeant Rabel did act, falling on the grenade and sacrificing his own life to save those of his companions.

It is not diminishing Sergeant Rabel's heroism to say that The Volunteer's Dilemma is not usually that extreme. It can amount to no more than offering to put the trash out. But how can we make the decision?

### 5.5.1 Credible Commitment and the Volunteer's Dilemma

The best answer is for people to *want* to take action on behalf of the group, even though it involves some sacrifice (small or large) to themselves. Offering unconditional positive regard may be seen as such a sacrifice, at least in the short term. So might taking action to relieve others of the responsibility. In both cases, the

conditions that have repeatedly appeared in the discussion of other dilemmas come into play—in particular, the spontaneous evolution of trust within small groups.

## ***5.6 The Battle of the Sexes***

At last, a problem for which game theory has a solution! It is a problem that confronted me and my English wife when we agreed to divide our time between Australia (where I was born) and England. The problem was that I would like to spend more time in Australia, she would like to spend more time in England, but both of us would rather be together than apart.

The answer was discovered by the Israeli-American game theorist Robert Aumann, who shared the 2005 Nobel Memorial Prize in Economics “for having enhanced our understanding of conflict and cooperation through game-theory analysis”. Aumann’s answer was for both people to agree to some random way of determining their strategy, such as tossing a coin or drawing a card. In our case, it was the toss of a coin, with the prearrangement that if it came up “heads” she was to stay longer in England before coming out to Australia to join me, with the reverse arrangement if it came up “tails”.

We were both better off with this arrangement. Aumann called it a “correlated equilibrium”, because it binds the choices of the two parties together in a very neat way. It may seem trivial when a coin toss decides the issue, but Aumann has proved mathematically that it is the most efficient strategy. It can even help to resolve some games of “chicken” where the participants seem to be locked into a mutually destructive collision course, with neither prepared to give way.

### **5.6.1 Credible Commitment and the Battle of the Sexes**

The only problem is credible commitment to abide by the result of the coin toss. Mutual trust is essential. Answers on a postcard, please (they can be copied from any of the discussions above).

## ***5.7 Stag Hunt***

The name of this dilemma comes from a story told by the French philosopher Jean-Jacques Rousseau (1754) about a group of villagers hunting a deer:

If a deer was to be taken, every one saw that, in order to succeed, he must abide faithfully by his post: but if a hare happened to come within the reach of any one of them, it is not to be doubted that he pursued it without scruple, and, having seized his prey, cared very little, if by so doing he caused his companions to miss theirs (para. 9).

Rousseau saw the story as a metaphor for the eternal tension between social cooperation and individual freedom. In his words, when referring to the “social contract” between the individual and the state: “True freedom consists in giving up some of our freedoms so that we may have freedom”.

Stag Hunt represents the fragile circumstances in which so many of the world’s people now live, especially when it comes to the preservation of individual liberties, freedom of expression, and even the freedom to hold private conversations. When I visited Tibet recently, for example, I found that it was impossible to talk freely with individual Tibetans about the problems in their country because they were frightened that their conversations, or even the fact that they had had a conversation with a Westerner, would be reported by one of their neighbours to the authorities. The Stag was the freedom to talk. The Hare was the more certain reward of spying and reporting secretly on your neighbour. Divide-and-rule works. It is not an easy thing to change, even with the tools of game theory.

### 5.7.1 Credible Commitment and Stag Hunt

I do not pretend to have an answer, even a theoretical one. If I had, I would be out there, shouting it from the rooftops. To solve such problems requires credible commitment and trust on a massive scale. Perhaps, sadly, the human race is simply not ready for it.

## 6 Fairness and Empathy

Game theory provides an accurate description of what happens *if* we rely on the logic of self-interest to guide our actions and our interactions. Our very human feelings for fairness and empathy, however, can turn this logic on its head.

One example, which surprised game theorists and psychologists alike when it was first observed, occurs in the “ultimatum game”. This game has been played primarily in psychological laboratories (usually with students as subjects) although it has many uncomfortable parallels in real life.

In the game, an experimenter gives an amount of money or other goods to someone, who is then required to offer a proportion to a second person. The second person can then either accept or reject the offer. If they accept it, the money or goods are shared accordingly. If they reject it, neither of them gets anything. That’s it. There is no further bargaining; it’s a one-off.

What should the “proposer” do? His or her obvious and logical course is to offer as little as possible, because the receiver has to accept it or get nothing. This sort of “take it or leave it” negotiating tactic has been widely used by the powerful to take advantage of the weak and helpless. It is a weapon for those in positions of power.

When researchers handed that power to volunteers in the “ultimatum game”, though, they received a surprise. They found that most “proposers” did not try to



keep as much as possible for themselves, but offered around half of the total, even when real money was involved. Even more surprisingly, when “receivers” were offered less than 30 %, they often exerted their own power by rejecting the offer, even though this meant that they lost out along with the proposer. “Receivers” seemed very willing to cut off their noses to spite the other person’s face—not only in affluent America, but also in countries such as Indonesia, where the sum to be divided was a hundred dollars, and where offers of thirty dollars or less were frequently rejected, even though this was equivalent to 2-week wages (Cameron 1995)!

Our inbuilt senses of empathy, and the altruism to which it can lead, can also help. These senses can be swamped by the perceived need for self-preservation in today’s often anonymous and depersonalised society, but they are always there, even in toddlers and chimpanzees (Warneken et al. 2007). There is some evidence that altruistic behaviour provides us with a physiological reward (in the form of the release of brain chemicals that give us the “warm glow” (Moll et al. 2006; Tankersley et al. 2007; Harbaugh et al. 2007)).<sup>7</sup> It would be absurd reductionism to say that brain chemistry and physiology alone account for our feelings, but they obviously play a substantial part. Whatever the origin of these feelings, though, it is clear from our earlier discussion that they offer the most substantial hope of overcoming the serious social dilemmas exposed by game theory—so long as we can learn to use them to create and maintain an atmosphere of mutual trust.

Such an atmosphere arises in encounter groups, where sharing works on two levels. One has its basis in game theory. If I share a personal secret with you, this makes it safer for you to share a secret with me, because you know that, if I betray your secret, you are in a position to betray mine.

The other level is psychological and lies in the fact that sharing creates an empathic bond. There is also a run-on, positive feedback effect that can run through the group, with one bit of sharing triggering memories for others and enabling them to share similar experiences. As those who have felt it know, the empathic bonds that are thus created can be extraordinarily powerful. As Professor Renate Motschnig pointed out in her review of an earlier version of this chapter, the next step is to understand further what “‘causes the magic’, how it could be transferred to the ‘real world’ and how aspects of it, combined with other PCA characteristics, could address the dilemmas”.

## 7 Conclusions

The game theorist’s approach to credible commitment lies in logic-based strategies where the person offering the commitment demonstrates publicly and conclusively to the other parties that he or she would lose out if they went back on their word. An approach based on Rogers’ research and ideas offers a different route—the

---

<sup>7</sup> This “reward” motive provides an interesting link between psychology and game theory, as I discuss in detail in *Rock, Paper, Scissors* (2008).

development of genuine trust through personal interactions based on unconditional positive regard, or in group situations where trust can develop spontaneously.

Both of these approaches have their place when it comes to resolving the problems posed by The Seven Deadly Dilemmas. They can work more effectively (especially those based on Rogers' research and ideas) in societies that contain many small groups, rather than consisting of a large and relatively homogeneous mass. The important thing in either case, though, is that we should be aware of these dilemmas, and of the serious social threats that they pose. Only then can we make genuine progress towards their solution.

## References

- Allesina, S., & Tang, S. (2012). Stability criteria for complex ecosystems. arXiv:1105.2071v1.
- Anon, J. (2005). Homo economicus? Sound economics may lie at the heart of humanity's evolutionary success. *The Economist* (April 7th). (<http://www.economist.com/node/3839749>).
- Berne, E. (1964). *Games people play*. New York: Ballantine Books.
- Cameron, L. (1995). Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. Working paper #345, industrial relations section, Princeton University. In R. Slonim & A. Roth (1998) Learning in high stakes ultimatum games: An experiment in the Slovak Republic. *Econometrica*, 66, 569–596.
- Fisher, L. R. (2008). *Rock, paper, scissors: game theory in everyday life*. New York: Basic Books.
- Harbaugh, W., Mayr, U., & Burghart, D. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, 316, 1622–1625.
- Hobbes, T. (Thomas Hobbes of Malmesbury). (1651). *Leviathan, or the matter, forme, and power of a common-wealth ecclesiasticall and civill*. London: Andrew Crooke, at the Green Dragon in St Paul's Church-yard. <http://socserv.mcmaster.ca/econ/ugcm/3113/hobbes/Leviathan.pdf>.
- Kuhn, H., et al. (1994). The work of John Nash in game theory *nobel seminar* (December 8). [http://www.nobelprize.org/nobel\\_prizes/economics/laureates/1994/nash-lecture.html](http://www.nobelprize.org/nobel_prizes/economics/laureates/1994/nash-lecture.html).
- Machiavelli, N. (1532). *The prince* (W. K. Marriott, Trans.). Chapter XVIII. (<http://ebooks.adelaide.edu.au/m/machiavelli/niccolo/m149p/chapter18.html>). The actual quote is "a wise lord cannot, nor ought he to, keep faith when such observance may be turned against him."
- May, R., Levin, S., & Sugihara, G. (2008). Ecology for bankers. *Nature*, 451, 893–895.
- Miller, D. T. (1999). The norm of self-interest. *American Psychologist*, 54, 1053–1060.
- Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences of the U.S.* (Vol. 103, pp. 15623–15628).
- Nash, J. (1951). *Non-cooperative games*. *Annals of Mathematics* (Vol. 54, pp. 286–294). This is surely one of the shortest papers ever to have won its author a Nobel Prize (the 1994 Nobel Memorial Prize in Economics).
- Ostrom, E. (2009). A general framework for analyzing sustainability of social-ecological systems. *Science*, 325, 419–422.
- Poundstone, W. (1993). *Prisoner's dilemma*. New York: Anchor.
- Rand, A. (1964). *The virtue of selfishness*. New York: Signet. See also Rand, A. (1992). *Atlas Shrugged* (35th anniversary edition). New York: Dutton.
- Rogers, C. (1942). *Counseling and psychotherapy*. New York: Houghton Mifflin.
- Rogers, C. (1951). *Client-centered therapy*. London: Constable.
- Rogers, C. (1961). This is me. In *On becoming a person* (p. 20). London: Constable.

- Rogers, C. (1970). *Encounter groups*. London: Penguin.
- Rousseau, J. (1754). *A discourse on a subject proposed by the Academy of Dijon: What is the origin of inequality among men, and is it authorised by natural law?* (G. D. H. Cole, Trans.). Rendered into HTML and text by Jon Roland of the constitution society. (<http://www.constitution.org/jjr/ineq.txt>).
- Russell, B. (1959). *Common sense and nuclear warfare*. London: Simon and Schuster.
- Smith, L. (1996). *Daughters of darkness*. New York: Simon and Schuster (This is the second novel in her *Nightworld* series).
- Tankersley, D., Stowe, C., & Huettel, S. (2007). Altruism is associated with an increased neural response to agency. *Nature Neuroscience*, *10*, 150–151.
- von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behaviour*. Princeton: Princeton University Press.
- Warneken, F., Hare, B., Melis, A., Hanus, D., & Tomasello, M. (2007). Spontaneous Altruism by Chimpanzees and young children. *PLoS Biology*, *5*, e184.
- Wilkins, P. (2010). Unconditional positive regard reconsidered. *British Journal of Guidance and Counselling*, *28*, 23–36.