

I Setting the Scene

There is a huge multidisciplinary literature on the significance of reciprocity to the functioning of societies, some of it dating back many hundreds of years. Indeed, to the extent that reciprocity can be thought of an exhortation to treat others as you wish to be treated yourself, the importance of this so-called golden rule as an ethical imperative is as old as the history of recorded thought, and is central to many of the world's major religions. In Luke 6.31 of the New Testament, for instance, Jesus asks us to do unto others as we would have them do unto us, and the Jewish religious leader, Hillel the Elder, alive at around the same time as Christ, reportedly went as far as to say that the golden rule is the *whole* of the Torah. In his classic book, *The Gift*, the social anthropologist Marcel Mauss (1954) noted that the Latin *do ut des* and the Sanskrit *dadami se, dehi me*, which both can be translated to 'I give in order that you may give', are found in Western and Eastern religious texts, and in the quasi-religious *Analects of Confucius* we are told that what we do not wish for ourselves we should not wish for others. Much of the above alludes, if sometimes loosely, to what is now known as reciprocal altruism or conditional co-operation; i.e., to repay kindness with kindness, and be willing to incur a cost in the expectation that it will be repaid in kind. Negative reciprocity, in the form of punishing harmful acts, akin to taking an eye for an eye, is embedded in the Old and New Testaments (e.g. Exodus and Matthew), the Torah, the Qur'an and, as we will see, in many more recent sources besides. Some may contend that the golden rule is too idealistic to be taken seriously in practice, but according to the evolutionary biologist Robert Trivers (1971), the notion that underlies the rule – namely, reciprocity – is observed in all known cultures.

2 SETTING THE SCENE

Simple acts of reciprocity are of course older than recorded thought and, since they are common in the animal kingdom, we can state with some confidence, predate humans. The reader will know, for instance, that cats lick each other and primates, literally and metaphorically, scratch each other's backs. This type of instinctive, attitudinal reciprocity is common among humans and between species also; for example, a person who has a door held open for her might unthinkingly hold the next door ajar in return, and a dog who licks its owner's hand may receive a pat on the head. There is not much, if any, pre-calculated deliberation involved in these acts. They are a type of direct reciprocity (Nowak, 2011), which refers to some kind of to and fro between two partners in a dyadic – i.e., a one to one – exchange. Tit for tat is also a form of direct reciprocity – i.e., if you help me, I will help you, and if you stop helping me, I will stop helping you.¹ The political scientist Robert Axelrod (1984) once famously – at least in social science circles – reported that tit for tat is the optimum strategy for human co-operation, although others have since concluded that it is a poor strategy when the information about the payoffs or behaviour of one's partner is less than perfect, or when the group is large and the number of people willing to co-operate is small (Henrich and Henrich, 2007).² Leaving this aside, attitudinal reciprocity appears

¹ Nowak (2011) also refers to what might often be a more calculated form of tit for tat that he defines as generous tit for tat, in which a person never forgets a good turn and occasionally forgives a bad turn. Also, since standard tit for tat is vulnerable to people misunderstanding others' intentions, which could lead to endless rounds of retribution, Boyd (1989) and Novak and Sigmund (1992) propose contrite tit for tat, where one side retaliates only if the other side has defected in two consecutive rounds, and forgiving tit for tat, where one side automatically forgives one third of the defections committed by the other side. On the flipside, in order to discourage defections from live-and-let-live truces between British and German soldiers during, for instance, mealtimes in the trenches of World War I, it was known on both sides that any rogue shell would be met by two in response (Sapolsky, 2017). These truces demonstrate that reciprocity can hold between enemies in the most extreme circumstances.

² Talcott Parsons (1951) noted, perhaps somewhat optimistically, that once a stable relationship of mutual benefit – akin to tit for tat, or rather tat for tat – has been established, it is self-perpetuating, and no additional mechanisms are needed to reinforce it.

to have arisen earlier in the evolutionary chain than a more deliberative form of reciprocity, conferring some form of advantage to those animals that behaved as such. In humans, attitudinal reciprocity may have served as the kernel for the later tendency towards more complex, co-operative endeavours.

So far, then, I have referred to reciprocal altruism, negative reciprocity, attitudinal reciprocity and direct reciprocity, and when one delves into the rich multidisciplinary literature on this topic, these and other definitions of reciprocity, sometimes with more than one term referring to the same concept, frequently arise. Given this confusion of terminology, it seems apt to spend a little time outlining the main concepts of reciprocity, before honing in on those that will inform the rest of this book.

TYPES OF RECIPROCITY

In a book published in 2008, the French economist Serge-Christophe Kolm wrote:

Reciprocity is to treat other people as other people treat you, voluntarily and not as a result of a binding exchange agreement . . . Good social relations in general, which are essential for the amenity and value of a society, are sustained by reciprocity . . . It is no surprise, therefore, that most social reformers have advocated a greater role for reciprocity as the alternative to coercive hierarchy, selfish exchange, and the utopia of unconditional altruism. (Kolm, 2008, pp. 1–2)

One could contend quite reasonably that coercive hierarchy, in the form of command and control policies, selfish exchange, informed by standard economic theory, and unconditional altruism, in informing the creation and development of many welfare states in the post-World-War-II era, have had a powerful influence on public policy design, and that the 'greater role for reciprocity' has generally gone somewhat unheeded. Kolm, to emphasise the richness of reciprocity, goes on to classify this notion into the following three types, all of

4 SETTING THE SCENE

which go beyond mere instinctive attitudinal reciprocity and rely more on deliberation and stored memory:

1. **Balance or Matching Reciprocity.** The motivation here, according to Kolm, is to return a gift with the aim of restoring some property of balance or fairness between the exchange partners.
2. **Liking Reciprocity.** Here, the return gift is motivated by a positive affective sentiment towards the initial giver, a feeling that may arise when the initial giver's gift or favour is perceived to have been offered benevolently. Kolm maintains that the sentiments underlying liking reciprocity are those of altruism and affection, and it is plausible that the relationship embedded in this type of exchange matters over and above the actual goods exchanged, or, in other words, the exchange has an intrinsic and not merely instrumental value.
3. **Continuation Reciprocity.** This is akin to exchange driven by self-interest, whereby all partners benefit by continuous positive reciprocal acts, and with retaliation to deter unfavourable actions – which Trivers (1971) termed moralistic aggression – as its negative counterpart.³ If the number of people who are willing to punish transgressors in a particular population is too small, egoistic actions become more profitable and may drive out reciprocal altruism.⁴ Many might see continuation reciprocity as the

³ Some people are willing to pay a personal cost to punish transgressors and are hence called altruistic punishers, but given that neural imaging studies have shown that people often appear to gain pleasure in punishing and taking revenge (Gospic et al., 2011; Strobel et al., 2011), perhaps these punishers are not always as altruistic as they might on the face of it appear. Moreover, although incurring an immediate cost, the possibility that altruistic punishers expect their actions to reap longer term gains for themselves cannot be discounted (Trivers, 1971). Nowak (2011), who tends to believe that punishment is counterproductive, prefers to call this costly rather than altruistic punishment, partly because he believes that the punisher can, in fact, bear some cost to themselves and partly because he takes the view that punishment in real life is hardly ever altruistic – that punishment can mark an escalation in conflict and is often motivated by anger or greed, or to suppress or exploit others. Sapolsky (2017) concurs, believing that punishment should never be considered a virtue.

⁴ The sociologist Alvin Gouldner (1960, p. 173) defined egoism as 'a salient (but not exclusive) concern with the satisfaction of one's own needs'. He therefore sees reciprocity, when driven by long-term mutual self-interest, as being egoistical, writing that 'the norm of reciprocity enlist[s] egoistic motivations in the service of social system stability'. However, in this book, egoism will be taken to mean *the exclusive* and somewhat shortsighted concern with the satisfaction of one's own wants and/or needs. It can be thought of as a highly selfish form of self-interest, with an egoistic actor willing to transgress group dynamics to benefit herself entirely whenever the

principle motivation for the reciprocal altruism and negative reciprocity mentioned earlier – i.e., it is in the self-interest of all exchange partners that these actions occur. This was the view taken by the Scottish Enlightenment philosopher David Hume, who in his *A Treatise of Human Nature* (1740) saw reciprocity as being motivated not by kindness, but because one expects the other person to return in kind in the future. Kolm himself, however, unlike the fairness and altruism embedded in balance and liking reciprocities, does not believe that actions driven by self-interest are pure forms of reciprocity.

Putting to one side Kolm's assertion that continuation reciprocity is not in fact a pure form of reciprocity, he does recognise that the three motivations defined above coexist across people, and often within a single individual. Moreover, he alludes to two further types of reciprocity, which are less about fundamental initial motivations of fairness, altruism and self-interest in a direct dyadic exchange, and are instead focused upon how people's reactions to these initial exchanges might have knock-on effects across a wider community:

4. **Generalised Reciprocity.** This is the notion that having benefited from somebody's help, a person is more likely to help other people, and is defined as indirect reciprocity by Sapolsky (2017). We can transport ourselves to a ready example of indirect reciprocity, whereby a car driver, on being let into the flow of traffic by another driver, will, one would hope, be more likely to give way to other users of the road. Thus, indirect reciprocity can involve a large number of actors, where, rather than the dyadic relationship in direct reciprocity, the individual's relationship is that between themselves and society in general, and thus this type of reciprocity affords wider co-operation within a group. We will see later in this book that fostering indirect reciprocity ought to be an important consideration in public policy.

opportunity arises, and therefore differs from the enlightened self-interest that often drives reciprocal relationships. It is typically (if, admittedly, not necessarily correctly) taken that the implicit assumption in standard economic theory is that human self-interest is egoistic rather than enlightened, and much of the time, enlightened self-interest will require somewhat more sophisticated – if often still almost instinctive – strategic decision-making and lower rates of time preference (i.e. a tendency to be patient) than selfish egoism. Also, as an aside but as will be discussed in more depth in Chapter 9, it is assumed here that the objective of an individual's enlightened self-interest is decided by the individual herself, and may differ across individuals.

6 SETTING THE SCENE

5. **Reverse Reciprocity.** Here, Kolm argues, people who are known to help others are themselves helped by others still, because they are thought of as deserving, a phenomena that is also often known as indirect reciprocity (and thus for the purposes of this book, acts of generalised reciprocity and reverse reciprocity will be both labelled as indirect reciprocity). A person's reputation is therefore fundamental to the operation of reverse reciprocity, which is an important reason why reputation has evolved to be such an important part of most people's psyche. To paraphrase Napoleon, men will die for a bit of ribbon to be placed on their chests, or to quote directly the biologist Martin Nowak (2011, p. 54):

Thanks to the power of reputation, we help others without expecting an immediate return. If, thanks to endless chat and intrigue, the world knows that you are a good, charitable guy, then you boost your chance of being helped by someone else at some future date. The converse is also the case. I am less likely to get my back scratched, in the form of a favour, if it becomes known that I never scratch anybody else's. Indirect reciprocity now means something like 'If I scratch your back, my good example will encourage others to do the same and, with luck, someone will scratch mine'.

Reputation serves as a proxy for experience⁵; Nowak (p. 67), quoting the evolutionary biologist David Haig, goes on to state, 'For direct reciprocity you need a face. For indirect reciprocity you need a name'. Ebay and Amazon sellers – to name just two domains – rely on publicised feedback to generate a positive reputation, in the hope that buyers will 'co-operate' with them, and if good reputations spread quickly enough, co-operation in a society has a better chance of

⁵ This notion applies to goods and services as well as people. For instance, there is a small noodle shop in central London called Kanada-Ya, that always has a large number of people outside it queuing for a table. I have no idea if the food is any good, but the queue signifies that it must be, and on passing it I always want to taste the noodles that it has to offer (although not so much as to join the very long queue). Similarly, the scarcity value that people attach to goods that appear to be running out – e.g., 'only two left in stock' – may be underpinned by the thought that, although we do not know it directly, the apparent high demand for the product indicates that it must be good.

emerging and continuing. I shall return to the importance of reputation to reciprocity and co-operation at several points in this book. As noted in the quote from Nowak, indirect reciprocity (and for that matter all of the other types of reciprocity mentioned above), can also have a negative counterpart. That is, a person who feels maligned by one person may also be more likely to malign other people – think of the car drivers – and a person who is unhelpful to others may garner a poor reputation and, as a consequence, may be shunned – or worse – by one and all.

As stated above, Kolm acknowledged that the fundamental motivations for acting reciprocally – i.e., for the sake of fairness, altruism and self-interest – can coexist, and indeed may intertwine (e.g. the notion of a fair exchange to promote the self-interest of all trading partners). It is very difficult, if not impossible, to tease out the relative impacts of these types of motivation on reciprocal behaviours, but one of the most repeated and robust experiments in the field of behavioural economics – the so-called ultimatum game – has uncovered reciprocal patterns in respondent groups in many countries that can plausibly be attributed to at least some of the types of reciprocity outlined above. As a precursor to various discussions offered later in the book, and as an illustration of how some of the motivations outlined above might drive people's choices in simple controlled contexts, it may not be without merit for the reader to be given an ultimatum.

A SIMPLE EXPERIMENT

In the ultimatum game, first reported by the economist Werner Güth and his colleagues (1982), each participant is assigned as either a 'proposer' or a 'responder'. Proposers and responders are paired and typically remain unaware of one another's identity. The proposer is given an amount of money, which can be any amount but is often in the range of £10–20 (or \$), and is asked for the share of her endowment that she is willing to give to her responder. The proposer is told that if the responder accepts the share offered, then the proposer is left with

8 SETTING THE SCENE

the remainder, but if the responder declines then both parties receive nothing.

The argument is often proposed that, according to standard economic theory, with its underlying assumption that people are invariably entirely selfishly, or egoistically, self-interested, the proposer should offer a tiny percentage of her endowment, because she will want to keep as much of the money as possible, and responders will accept tiny offers, because they will consider anything at all to be better than nothing. However, the general finding across multiple studies is for the mean offer from proposers to exceed 40 per cent of their endowment, for modal offers to be 50 per cent, for offers of less than 30 per cent to be frequently rejected by responders and for offers of less than 20 per cent, although rare, to be rejected half of the time (for reviews see Camerer and Thaler, 1995; Henrich et al., 2005; Kahneman, Knetsch and Thaler, 1986; Roth, 1995).

Some economists, including Kenneth Binmore, maintain that these results do not, in fact, necessarily undermine the descriptive validity of standard economic theory, by arguing that it is not axiomatic in economics that utility increases with money holdings, and that whether utility maximisation equates to income maximisation is an empirical question (Binmore, 2005). They rightly claim that the experimental approach to understanding social norms is common among game theorists, that standard theory is more nuanced than critics often give it credit for and that standard theory requires only that people are consistent in their choices. Behavioural economists and others have amply demonstrated that this latter assumption is, at the very least, heroic, but it is true that standard theory merely requires that people maximise *something* (i.e. not necessarily money). Economists could of course raise many other objections to the findings of the ultimatum game, not least in relation to its external validity. For example, it is uncommon in real life that people are given what may be viewed as essentially 'free' money (i.e. windfall money), and proposers and responders might both react differently to the figures summarised in the previous paragraph in the more realistic

context of proposers being required to do something to earn their initial money holdings.⁶

Nonetheless, it can reasonably be contended that the ultimatum game is a simple device that demonstrates that people are influenced by different reciprocal motivations in particular contexts; that man is less *homo economicus* in the common understanding of the term, and more *homo reciprocus*, in the words of the sociologist Howard Becker (1956). For instance, many proposers in the typical ultimatum game are motivated to offer a substantial share of their endowment by the desire to maximise their final holdings, given their knowledge that if they offer an amount that is deemed unfair by the responder, they will end up with nothing.⁷ That is to say, that they are aware that responders, implicitly driven by balance reciprocity, are motivated to respond negatively to very low offers, and thus the proposers, driven by something resembling continuation reciprocity, offer an amount that they feel will be just sufficient to avoid punishment, thereby – in their mind – maximising their expected final holdings.⁸ If this occurs, the proposer is therefore fundamentally self-interested, but is sensitive to the possible reactions of other human beings to her decisions.

⁶ In recent unpublished experiments that I have conducted that replicate the ultimatum game with the exception that proposers have to undertake tasks to earn their initial endowments (and with responders being aware of this fact), proposers tend to offer shares of their endowments that are somewhat smaller – although still substantial – than that generally observed, and responders are typically willing to accept nothing (or very close to nothing). The responders' reactions were motivated by something akin to balance reciprocity; i.e., they took the view that since the proposers had earned and thus deserved the money and that they themselves had done nothing, it would be unfair to punish the proposers for low offers.

⁷ Gintis et al. (2005) report a variant of the ultimatum game where rejection by responders left only them, and not the proposers, with nothing, which led to responders never rejecting offers, suggesting that rejection is driven entirely by the desire to punish. In this variant, proposers made much smaller but still positive offers, implying a residual of fairness amid the self-interest.

⁸ The ultimatum game is often a one-shot game, and thus in these circumstances contrary to what the 'continuation' in continuation reciprocity implies, the proposers and responders are not forging a longstanding relationship based on mutual self-interest. However, they are attempting to forge an arrangement that leads, or continues, to a final outcome where they receive allocations of the endowment that both find acceptable.

10 SETTING THE SCENE

As earlier noted, a condition of anonymity between the proposer and responder is usually assumed in the ultimatum game, but in many social and economic exchanges the parties are aware of each other's identity. If the participants knew each other in the ultimatum game, we might expect liking (or disliking) reciprocity to play a role, perhaps motivating the proposer to offer more of the endowment, and/or for their offer to be motivated by affection rather than self-interest, and for the responder to be slower to punish. To a certain extent, liking reciprocity can also perhaps be biologically manipulated, and in non-anonymised circumstances, strengthened in the ultimatum game. For example, the neuroeconomist Paul Zak exposed proposers in a version of the ultimatum game to a nasal spray containing oxytocin (Zak, 2012). Oxytocin is the hormone that prepares female mammals for childbirth and lactation, helps mothers to bond with their offspring and is often labelled as a facilitator of trust. Zak reported that this increased offers by 80 per cent compared to placebo. He further reported that women, who as one might expect naturally produce more oxytocin than men, tend to be more generous in these somewhat abstract games, although Henrich et al. (2005) did not observe any strong or consistent effect of sex on ultimatum game responses. Zak also noted that infusions of testosterone – which blocks the binding of oxytocin to its receptors – was associated with ultimatum game offers that were 27 per cent less generous compared to placebo. Before concluding that oxytocin should be released into the water supply, however, the neuroendocrinologist Robert Sapolsky (2017) warns that this hormone has a dark side. For instance, when the players in cooperation games remain anonymous to each other and are placed in different rooms, oxytocin *diminishes* generosity. It therefore appears that oxytocin strengthens the bond with those you know, but turns you against those who, in an evolutionary sense and particularly at times of childbirth and rearing, you might perceive as a threat. Furthermore, in the right circumstances, testosterone has a pro-social bright side. For instance, the neurobiologist Christoph Eisenegger, with colleagues including the economist Ernst Fehr, reported that