Reciprocal Altruism in the Theory of Money Daniel Krawisz

nakamotoinstitute.org/reciprocal-altruism-in-the-theory-of-money/

December 8, 2014

Methodological Issues

Richard Dawkins said in an offhand comment in *The Selfish Gene* that "Money is a formal token of delayed reciprocal altruism."[i] This turns out to be a rather insightful way of looking at money, and the purpose of this essay is to explore the idea more deeply to see how far it can take us. Nick Szabo later used some of the ideas in *The Selfish Gene* to describe the historical origins of money in his essay "Shelling Out"[ii], but this essay will be about the theory of money.

It is first necessary to get some methodological issues out of the way. Biology and economics are similar in the way that they treat the interactions of many individuals in terms of the incentives that they all put upon one another. They then find the strategies which are most successful under the circumstances. In biology, especially in the theory of social evolution, this is often treated explicitly in terms of the language of game theory[iii]. In economics theory this is done somewhat less often, but ultimately any discussion of incentives, which is what economics consists of, can be treated in terms of game theory. Both theories suppose that the strategy which produces the greatest benefit for individual actors will tend to win out. In biology, it is assumed that the natural selection is the means by which this happens, whereas in economics this happens because of learning or cultural evolution.

Individualism is important in both biology and economics. In biology, the problem is often to explain how highly cooperative and altruistic behavior can be explained in terms of the self-interested behavior of organisms attempting to spread their genes. In the 1960's, an idea had become popular in biology called group selection, which is that selection can act on groups of organisms rather than just on individuals, and consequently that organisms can have traits which can be deleterious to individuals yet beneficial to the group as a whole. However, if any individual in a group could out breed its fellows by reducing those deleterious traits, it would. Adaptations which appear to be good for the group must therefore, always be good for the individual[iv].

In economics it is the other way around: the problem is often to show how certain kinds of rules—say a government regulation—produce adverse effects because they fail to be individually beneficial to the people subject to them. The greater the resources that the government employs enforcing its rules, the greater the benefits to successful cheaters. Therefore, no amount of expenditure is ever enough to produce the desired outcome.

Both sciences search for functional explanations of behavior in terms of rewards and punishments rather than in terms of mental processes. In biology, an animal's psychology is just like any other body part—it evolved to serve specific purposes. Therefore, we do not explain animal behavior in terms of how it feels or wants. Its feelings and desires are to be explained in the same way as everything else about it: as part of a strategy to maximize expected future offspring. In economics, we wish to explain peoples' actions in terms of consumption. Consumption may be anything that may be treated as an immediate reward. It does not matter what a person claims about why he does something. What matters is what he actually chooses, and the problem of economics is to explain his choices in terms of his preferences. It is easy to see that this approach is necessary by the way people treat money: though everyone uses it every day, very few can give a satisfactory explanation of it.

Thus, both sciences are behaviorist in a certain sense. As in behavioral psychology, we are attempting to explain behavior in terms of rewards and punishments, rather than in terms of invisible feelings, behavioral psychology attempts to explain behavior as the result of a past schedule of rewards and punishments, whereas economics and evolutionary biology attempt to explain behavior in terms of future expected rewards and punishments. Experimental

psychology thus treats organisms more like physical processes which are expected to respond in predictable ways to past stimuli, whereas in economics and evolutionary biology, organisms are goal-oriented, machines that respond to incentives (that is, expected rewards and punishments) in search of some optimum.

There is a common criticism of economics which says that economic theory is invalidated because real people are irrational, whereas the people in the theory are much too rational. Biology is a good standpoint to explain the problem with this objection. Typically in game theory, we define equilibrium strategies in terms of players with perfect knowledge of the game, who are able to compare the likely outcomes of every strategy and who assume the same of their opponents. Of course, it is not true that the organisms being modeled by the game actually know anything about strategy—or anything at all—but the game-theoretic models still work. This is because the equilibrium strategies of the game (if they exist) are precisely what an adaptive system will tend toward, even if it is not intelligent or rational at all.

All life adapts by natural selection, but some, preeminently humans, adapt readily by learning. However, the mechanism of adapation is irrelevant: it just means that humans adapt much more quickly than other animals. The objection that humans are too irrational for economic theory to apply to them is a failure to think in behaviorist terms. People must necessarily adapt to a set of rewards and punishments, and they do so by learning. We do not need to assume that people reason their way to the best strategies in a game; we only need to assume that people tend to imitate the more successful among them, and that consequently the more successful behaviors tend to beat out the less successful. The people do not need to understand what they are doing or why in order for this process to work.

Furthermore, applications of economic theory normally depend on how the optimal behavior changes rather than its absolute value. This allows us to make statements about the economy without knowing precisely what the optimal behavior is and without assuming that anybody actually is behaving optimally. A typical economic prediction takes a form like the following: “circumstance AA rewards behavior xx more than circumstance BB. Therefore if there is a change from BB to AA, we should eventually expect more of behavior xx than before.” As long as the premise of the argument is true, then the prediction that there will be more behavior xx can be made with great confidence, but not with great precision. Because the precise way that people will react to a change from AA to BB depends on how strongly the two situations reward or punish them, and on how quickly they learn to adapt, it would be very difficult to say how much more xx could be expected. However, ceterus paribus, a change from AA to BB certainly cannot cause less xx. Thus, for example, I can say without hesitation that a minimum wage, if it is set high enough, will cause unemployment, but I cannot say how much.

Those are the similarities between biology and economics. The most important difference between them is the way that they treat value. In economics, value is subjective. People can value many things, even things that do not make sense, and it is not the economist’s job to ask why that should be. Goods which are considered to be valued for the immediate satisfaction they provide when they are consumed are called consumer goods. Not all goods are consumer goods, however, and the value of everything which is not a consumer good must be explained in terms of the consumption that it ultimately makes possible. For example, people generally do not want to own factories just because they like factories; they want to own a factory in order to earn a profit with it, which they can then spend on consumption. A factory is an example of a capital good. Ultimately its existence in the economy can be explained because it can be used to produce other goods.

On the other hand, in biology, value is not subjective. There is one ultimate value, and that is the maximization of an individual’s expected future offspring, or more properly, the maximization of the expected rate with which its genes will spread throughout a population. A puzzle in biology is when an organism exhibits behavior that appears to devalue things like food and status because it is very easy to see how such things would promote its survival and mating success. If an organism gives food away or rejects status, then that needs to be explained. On the other hand, in economics there is no particular need to explain someone who is an anorexic or a cynic philosopher. That would be a job for a psychologist; to an economist those are just his preferences. Consequently, whereas in economics, interpersonal comparisons of value are impossible, in biology they are possible between members of

the same species.

Concepts in biology can be carried over directly into economics by taking account the difference between the treatment of value in each. If a biological idea which can be adapted to explain economic behavior in terms of the consumption that it ultimately enables—accommodating the fact that people may have very different preferences about what they like—then it is as valid in economics as it is in biology.

Altruism in Economics and Biology

To the economist, altruism simply means gaining satisfaction from benefiting other people. Nothing about that requires an economic explanation. Peoples' preferences are exogenous, and if they enjoy helping others, that is not an economic issue per se. Whereas in biology, the only value is that which best spreads genes. It is easier to understand how this should produce selfishness on the part of an organism, so an organism which enjoyed helping others would be in need of an explanation.

In biology, altruism cannot be explained in terms of feelings and satisfaction because feelings are difficult to observe. Altruism needs to be defined in terms of observable behavior. Many animals are not capable of having feelings at all, but even very simple organisms are capable of cooperating. We can therefore abstract feelings away entirely. There is even a concept in biology called parasitic altruism, which means that an animal is a host to a parasite and is better off continuing to feed the parasite than attempting to remove it, which would be too expensive. More likely an animal in that circumstance would feel resentment, not benevolence.

Therefore, in biology, altruism is defined is an action which reduces an organism's own fitness and increases the fitness of another. The word fitness has to be clarified. Of course no behavior could survive in a population that does not help to maximize an organism's offspring. Fitness must be understood in an experimental sense. An organism's true fitness is not easily observable because it is, first of all, probabilistic, and second because even to estimate it would require observing an organism over its whole lifespan and beyond to learn how many children it raised compared to others of its species.

Instead, experimental fitness is hypothesized to relate to something more easily observable, such as resource-gathering capacity and a study is done to disprove the hypothesis. For example, if resource-gathering capacity was our hypothesized definition of fitness, then an altruistic animal would be one that gave resources away to another. To observe a population in which animals habitually gave resources to one another would prove that resource-gathering capacity alone does not truly measure their fitness. In biology, altruism is a bit of a loaded word—it does not stand for its face value, but rather for a behavior in need of an explanation.

Any altruism that is observed in biology implies that there are benefits which take a longer time to play out than that which is required to make the observation. This is the only way that an observation of apparent altruism can be reconciled with the theoretical requirement of explaining everything in terms of individual benefit. This does not mean that every single altruistic act must be calculated to produce some future benefit. The benefit may be probabilistic. For example, consider the case of a bunch of soldiers who go to war for their home country. They are behaving altruistically toward the people of their nation because they are putting themselves at risk in order to keep their friends at home safe. Let us say that most of these soldiers will die. Can their altruism still be explained in terms of future benefits? Yes, certainly. For some of the soldiers will survive and will return home to receive great honor and status, and it is not known who will survive beforehand. If the expected value of ones future status as a veteran offsets the probable cost of dying horribly in battle, then it can be expected that soldiers will want to go to war.

The Problem of Money

The first step for talking about money is to define it in behaviorist terms and to pinpoint what about that behavior is most in need of explanation. We don't think of money as a thing; instead we think of it as a behavior. We're not

humans living and working in our own economy anymore; we're biologists or economists observing the human species and theorizing about why monetary behavior is a successful strategy.

In economics, the interesting thing about money is that it is not consumed, just held. Even a miser who loathed to spend money and who simply hoarded it without any apparent plans to spend it is not treating it like a consumer good because if the money he was hoarding hyperinflated and became valueless, he would presumably throw it away. (We could imagine someone who just loves to collect money regardless of whether it can be spent, but that person would just be a coin collector, and in that case he would be treating the money as a consumer good. There would be nothing to explain in that case.)

Of course there is no need to explain why someone would spend money because that is when he gets resources for consumption, which improves his fitness. There is, furthermore, no reason to explain why someone would rob money at gunpoint or burglarize a house for money. If we know why someone would trade goods and services for it, then we know why someone would steal it. Another thing that does not need to be explained is why someone would manufacture or counterfeit money. None of this is altruistic.

The puzzling thing about money is that everyone wants it in the first place. If that can be explained, then everything in the previous paragraph is explained immediately. In economic terms, a person who accepts money in payment gives up a good that can be consumed for something he does not intend to consume. In biological terms, he makes himself less fit by taking the money and makes the buyer more fit because he gives up real resources or incurs a cost to himself in energy or time to the immediate benefit of another. From a biological standpoint, this is clearly an example of altruism, although not for the economist. However, from the standpoint of both sciences, the same behavior is in need of explanation.

This behavior I call monetary behavior, and I define it as the acceptance of money in trade; i.e. monetary behavior is to trade something for a good whose most valuable use to the one receiving it is to trade it again later. I wanted to coin the term monetary behavior because we tend to be so used to thinking reflexively of money as something that everyone wants and of the value of money being somehow in the money good itself that it is difficult to remind ourselves consistently that this is not the case. Thinking of money as a behavior rather than as a thing is a way of consciously reminding ourselves that the only possible value of money is other people.

It may seem obvious why monetary behavior exists: people want to spend it on something else later! However, there is a problem here: they only want money because other people also want money. Which is the same reason they all want money too! Everyone wants money because everyone else wants money because everyone else wants money… This is really pretty extraordinary. How can a whole society behave this way? It sounds like a it is holding itself up by its own bootstraps. One could object at this point that there is no infinite regression because originally what is now money may have been used for other purposes. However, recall that economics is future oriented, not past oriented, so the past uses of the money good are irrelevant. It matters not a whit to me whether a gold coin will one day be used to make a watch or necklace. As far as I care, it could well continue to be traded as cash forever. The problem is to explain how monetary behavior can be sustainable.

One might say that he is selling goods because he produced much more of a good than he would want himself, for the very purpose of trading it. Typically people specialize in what they produce and then use the money they make to buy what they need. However, that puts the cart before the horse. Specialization to that degree depends on the existence of money, not the other way around. This explains why money is socially beneficial, but we need to explain why it is individually beneficial. It has to be explained in terms that would make people want to start using it before they depend on the highly specialized economy that it enables.

As I observed above, altruism implies no particular emotional state. A moneymaker may not feel very altruistic about amassing wealth, but to convince everyone to use money would be a roughly similar problem to convincing everyone to live in a commune "from each according to his ability and to each according to his need." In a commune, I would depend on the good behavior of everyone else. If I worked hard and everyone else slacked off, then I would

be taken advantage of. Everyone must do his duty or the system fails. By the same token, if I begin to accept payment in money and to hold savings in it, I am depending on other people behaving the same way. What if I started saving in money, but no one else did so? Then I will have done a lot of work for nothing because my money would not be accepted anywhere. Thus, it is not just a rhetorical move to call money a form of altruism; there is a very deep and conceptually useful sense which makes the use of money very similar to other behaviors which people would more readily identify as altruistic.

Reciprocal Altruism in Simple Games

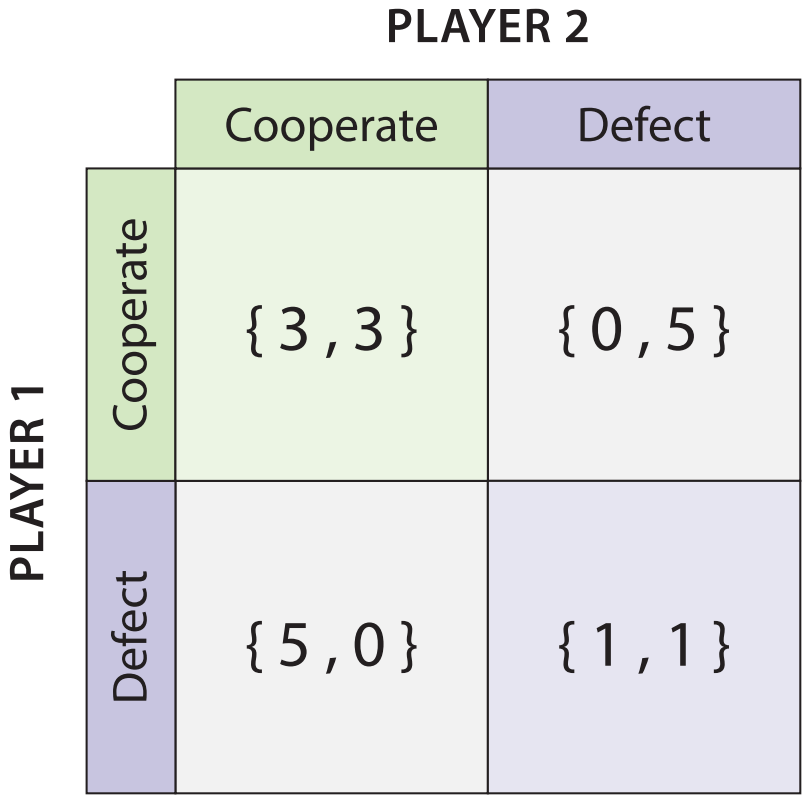
An allegory from many folklores says that hell is a table full of delicious food, but everyone must use such long utensils that no one can bring the food to his own mouth. Therefore, everyone starves in the midst of plenty. Whereas heaven is exactly the same, except that everyone feeds the people sitting across from him rather than attempting to feed himself. This is essentially the idea of reciprocal altruism, which is an idea introduced by Trivers in 1971[v].

The idea is that pairs of animals provide favors to one another and are better off because of the expected benefit of being able to receive favors when in need. Of course, if everyone were completely indiscriminate with their favors, no such system could persist because it would be open to abuse by cheaters. A selfish animal could benefit by receiving favors but never giving out any. Therefore the selfish behavior would be self-promoting, and would soon take over the population. In fact, as selfishness became more and more prevalent, altruism would become less and less beneficial because the altruists would meet one another less and less often, and therefore would spend more and more of their energy benefiting selfish animals. Unless there is a means to prevent cheating, reciprocal altruism fails to be individually beneficial. The problem in evolutionary theory is to clarify what makes this sort of interaction possible.

Theoretical studies of reciprocal altruism involve repeated two-person games. There are a variety of two-person games which allow for the development of reciprocal altruism. In these games there is one optimal strategy for a single round of the game, but a different optimal strategy if the game is going to be repeated many times. This is precisely what is expected; altruism must always be explained in terms of future benefits, so if there is no future, altruism is impossible.

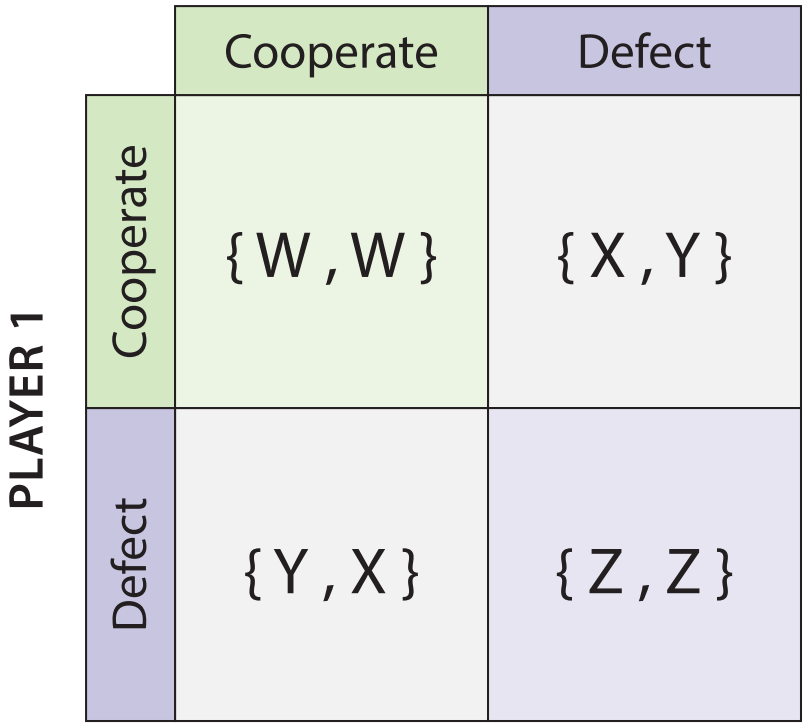
A successful altruistic strategy has two characteristics: first, it must be preferable to be an altruist with another altruist than to be a non-altruist with another non-altruist; second, players can choose how to react to the past behavior of their opponents. This allows a player to be altruistic with other altruists or non-altruistic with others who are not. Reciprocity discourages non-altruism. The non-altruist reaps what he sows. Under these circumstances, altruism will succeed in repeated games as long as the probability of repetition is high enough.

One example of a game that models reciprocal altruism is the famous Prisoner's Dilemma[vi]. This is a two-person game, in which each player has two options: cooperate or defect. The payoff matrix for the Prisoner's Dilemma looks like this:



In this diagram, each box represents an outcome after both players have made a choice. The first item in the list for each outcome is the reward to player one, and the second is the reward to player 2. The numbers in the boxes are arbitrary—what matters only is some ordering relations between them. A more abstract (but equivalent) payoff matrix for the Prisoner's Dilemma is this:

where Y>W>Z>XY>W>Z>X and X+Y<2WX+Y<2W. This second condition means that outcome WW is preferable to equal odds of outcome XX and YY.



The Prisoner's Dilemma may seem like an odd choice to model reciprocal altruism because both players must cooperate at the same time. That seems more like mutualism! However, each cooperative move on the part of either player is altruistic because, in the one-round case, every cooperative move is less beneficial to itself and more beneficial to the other player. Mutualism only occurs when it is immediately beneficial to both players to cooperate. It is possible to alter the game so that the players must alternate at behaving altruistic to one another, but it is easiest to understand the Prisoner's Dilemma first.

The one-round Prisoner's Dilemma is very simple—the outcome is that both players defect. This is the best outcome that either player can ensure for himself. They can't cooperate because neither can prevent the other from defecting. On the other hand, the iterated Prisoner's Dilemma is a much more complicated game because there are an infinite number of possible strategies. The game has still not been solved completely, so instead of attempting to find the best of all possible strategies, we enumerate a subset of simpler strategies and compare them. This is perfectly alright—normally we do not know for certain which strategies are the best in real-life scenarios. Instead we try to understand real situations by comparing a few alternative strategies. If we cannot, then we have to discover new strategies.

Before talking about specific strategies, however, we have to talk about how to evaluate them. Suppose two players cooperate every round, for an infinite number of times. They get a benefit of three each round, which comes to infinity. Now suppose they defect instead. That's a benefit of 1 each round—which also comes to infinity. How do we compare these outcomes? I know of two reasonable approaches, and both of which lead to the same qualitative results. One is to say that there is a cut-off and that the game ends after a fixed number of rounds. The other is to say that each round becomes successively less important the further in the future it is. In other words, a player who earns a reward XX each round for an infinite number of rounds earns

R3X+R2X+RX+X+…=X1−RR3X+R2X+RX+X+…=X1−R

where RR is some rate that determines how quickly the value of future rewards declines. This rate could be

interpreted as a probability, or as a rate of time preference, or as a combination of both.

- Always cooperate.
- Always defect.
- Tit-for-tat: the player cooperates on the first round and subsequently repeats whatever the opponent did on the previous round. This can be thought of as a strategy which punishes defection and rewards cooperation. Tit-for-tat is actually known to win out against a very complex array of strategies, not just the very simple ones considered here.
- Alternating: the player defects on the first move and subsequently alternates between cooperation and defection each turn. (This strategy is going to be important later on.)

Other simple strategies are possible, but they aren't interesting theoretically because they are neither successful nor do they make any intuitive sense either. I will not go through all the game theory mathematics but merely state some results[vii].

- Cooperation and tit-for-tat do equally well against one another.
- As long as no one plays tit-for-tat, defectors always do better than cooperators. This means that in a population of cooperators, if a single defector evolves, its strategy will propagate through the entire population until everyone is a defector.
- If RR is low enough, then defection succeeds no matter the composition of the population.
- On the other hand, if RR is high enough, then a single defector does poorly in a population of tit-for-tatters.
- A single tit-for-tatter does poorly in a population of defectors no matter how high RR is (it cannot be greater than one).

The gist of these results is that a population will always either end up all defectors or all tit-for-tatters, depending on RR and on the initial composition of the population. If tit-for-tat wins out, then it is indistinguishable from a population of cooperators because no one ever needs to be punished. Every round looks like pure altruism.

A crucial element of tit-for-tat is its charitable opening move. If one is the only altruist in a population of defectors, then it is very bad to give everyone the benefit of the doubt. For any proportion of altruists to defectors, there is always some interest rate such that it is better to be a defector, and the defectors do better and better as they become more numerous. Thus, with cooperation there is a network effect. It is better to be a cooperator among lots of other cooperators than among lots of defectors.

What exactly does this analysis prove? Rather a lot, actually. The Prisoner's Dilemma has a very wide applicability. Of course in real life, people generally have more options available to them that just cooperate or defect, but other options can be ignored if they do not affect the optimal strategy in a given circumstance. Thus, the Prisoner's Dilemma can apply to much more complicated circumstances than it may at first appear.

Furthermore, the game does not need to be played exactly the same way each round. The conditions that $Y>W>Z>XY>W>Z>X$ and $X+Y<2WX+Y<2W$ can be satisfied in many different ways by many kinds of possible future interactions. As long as the relations hold, it is not necessary to assume that each of WW, XX, YY, and ZZ refers to the same thing each round, or takes the same value to the players. It is still true that tit-for-tat can be the right strategy in any individual round as long as there is a high enough likelihood of similar interactions in the future.

So the Prisoner's Dilemma appears everywhere. To show more concretely how it can hide in other contexts, I will show how it appears in another two-person game. If the Prisoner's Dilemma is modified so that $X+Y>2WX+Y>2W$, then the players would be better off alternating cooperation while the other defected. This is called the modified Prisoner's Dilemma. This would seem more like a realistic model of reciprocal altruism, if we could get the players to

take turns with one another. In this case, tit-for-tat doesn't work as well because it can be beaten by the alternator. An alternator, however, is still beaten by a defector, so altruism does not win out with any of the simple strategies already presented.

It turns out the best strategy is one that mixes alternation with tit-for-tat, according to some probability whose value is determined by the precise nature of the payoffs. Two players following the mixed strategy will eventually become permanently in sync with one another. This strategy wins out if the probability of future interaction is high enough and it has a network effect is well. It retaliates against other players that go out of sync with it—eventually. Because it only follows tit-for-tat with some probability, it may take several rounds to react. If we allowed for more complex strategies, we could devise a strategy that retaliated more reliably. All it would require is for the player to react to the last two moves rather than the last one, but it would be very similar—it would be an alternator or a tit-for-tatter under a specific condition rather than a probability.

Of course, if more complex strategies were allowed, then the players could exchange favors according to more complex rules than mere alternation. One player could cooperate twice in a row as long as the other defected twice in a row. This would all work out as long as the players remained even on average. Of course, this would require a degree of coordination between the players which we have not allowed for. However, allowing for it does not change the nature of each individual round of the game, only the schedule of favors that may be available in the future. Thus, allowing for it does not change the conclusion that altruism is a successful strategy under the right circumstances.

So now we have two kinds of reciprocal altruism, right? Not really; there is nothing much new here because we can find the Prisoner's Dilemma hiding in them. This table explains how to combine moves in the modified Prisoner's Dilemma to construct the same outcomes as the standard Prisoner's Dilemma.

		PLAYER 2	
		Cooperate	Defect
PLAYER 1	Cooperate	Players cooperate and defect in sync	Player 2 defects twice in a row
	Defect	Player 1 defects twice in a row	Players mutually defect

Other possible moves exist in the modified Prisoner's Dilemma over two rounds, but they can be ignored because they do not affect the winning strategy.

Tit-for-tat is the winning strategy once again in this new way of looking at things, and that is what one is really doing with the winning strategy of the modified Prisoners Dilemma. Players still win by cooperating with other cooperative players and punishing non-cooperative players. It is just that cooperation plays out over several moves. Other two-person games with altruistic strategies can be treated the same way[viii]. There are a few different games which are similar to the Prisoner's Dilemma. Some of them allow for altruistic strategies and others do not. Those that do all have (at least) three similar properties. First, altruism can only succeed if the future is not discounted too rapidly, or alternately if the number of rounds of the game is high enough. In other words, the future must be important to the players of the game. Second, there is a network effect for altruistic strategies. Third, a successful altruistic strategy must be capable of retaliating against non-altruists.

Altruistic Groups

Obviously, the use of money does not require pairwise reciprocity, so it is quite different from the two-person games described in the last section. However, the two-person case is easy to generalize to larger groups. It is not logically necessary that altruism should always occur in pairs. For example, if animal AA was altruistic to animal BB, and animal BB was altruistic to animal CC, and animal CC was altruistic to animal AA, then their system ought to work out just fine, even though there is no pairwise reciprocation. This is the sort of thing that Dawkins apparently meant when he referred to "delayed" reciprocal altruism. There is not direct reciprocity between pairs of organisms, but the value of altruism is the same: the value of the favors that one eventually receives is worth the cost of giving them out. What goes around, comes around.

As with the iterated Prisoner's Dilemma, a complete theory of group altruism would be extremely difficult to elaborate, but we can still compare different kinds of simple strategies without describing every possible one, and the lessons from the two-person games hold true. For example, imagine an organism in a group that chooses to act altruistically to other organisms based on their past interactions with other group members. If organism AA watches organism BB behave altruistically toward organism CC, then, given the opportunity, AA will act altruistically toward BB. After that, AA may never interact with BB again, but other organisms may have seen AA's altruism toward BB and will act altruistically toward AA at the next opportunity. On the other hand, organisms in whom altruism is not observed do not receive it. Under circumstances in which the potential benefit of receiving favors is greater than the cost of giving them out, this sort of behavior will enable altruistic organisms extract that benefit without allowing non-altruists to mooch off of them.

Altruistic groups require more complex behavior on the part of the animals because detecting cheaters requires them to observe and keep track of more relationships than just their own. Nonetheless, some more complex systems of reciprocal altruism are known. For example, some very intelligent animals, such as apes, are able to form coalitions. They are more altruistic toward other group members than to non-members without always expecting direct reciprocity[ix]. Humans, of course, are very good at forming altruistic groups. I am sure that you have experienced this: when you learn that someone belongs to a group that you belong to, you feel much more willing to act altruistically toward him and vice versa, without reciprocity necessarily being expected from either side. This is possible because anyone can damage their reputation in the group by not acting altruistic within the group. (To be clear, I am not claiming to describe peoples' real feelings toward one another when they interact in groups. They are not necessarily going through a calculation like this. Rather, people just have an instinct to feel more warmth toward other members of groups to which they belong.)

It is possible to make the simple models of the previous section directly applicable to group cooperation by treating one player as a cooperative group that coordinates its treatment of another individual, represented by the other player. Suppose that NN organisms have learned to coordinate their treatment of one another. They can then encourage altruism in an N+1N+1th organism by treating it to a coordinated tit-for-tat strategy. When the N+1N+1th organism fails to be altruistic to any member of the group, the group as a whole must deny him the next time he needs help.

If the N+1N+1th proves to be capable of coordinating with the rest, then later they can open relations with an N+2N+2th organism in the same way. Group coordination can be considered to be a form of altruism in and of itself, so organism N+1N+1 is observed to follow the group rules with regard to organism N+2N+2, then it is rewarded or punished accordingly.

I'll go over this construction again because I have described several different interactions as two-player games and it is important to keep them all straight in under to understand why this should work. I have described how the organism N+1N+1 can encourage organism N+2N+2 to behave altruistically to the group of NN organisms. This is not any particular kind of altruism—just a reward or punishment that depends on whether organism N+2N+2 is altruistic or not. Of course, in reality, there would be many relationships of this kind. Every organism from 11 to NN would also treat organism N+2N+2 in this way. And they would also treat one another according to the same terms. I have also described how the rest of the group encourages organism N+1N+1 to coordinate with it. It is rewarded or punished according to whether it has correctly rewarded or punished organism N+2N+2. I have treated organisms N+1N+1 and N+2N+2 as if they were outside the group, but really they might as well be on the inside. Everyone in the group is also always treating everyone else according to the same rules, and every member is simultaneously playing simple games with one another and with the group as a whole.

Group altruism, therefore, can be individually beneficial under certain circumstances. If the members of a group have the ability and the incentive to keep track of their members so as to remember who has been the most altruistic, then altruistic groups can grow by accretion.

Money as Group Altruism

Money satisfies the conditions of reciprocal altruism and, in fact, allows for a much greater degree and extent than any other that has ever existed. It is so easy: everyone simply trades some kind of token with one another to keep track of how altruistic they have been. When you provide a favor to a group member, you get tokens from him, which you can later use to redeem favors from someone else in the group. There is no need to know anything about another person in order to detect cheating. You do not need to know his group loyalty. You only need to know whether he has money or not. If he has run out, then he needs to provide some favors rather than call them in. Cooperation with money connects all people in an economy, regardless of religion, nationality, or other group loyalties.

A money system does not place any practical limits on the size of the group. Anyone can enter the group of altruists by choosing to perform a favor in exchange for money. Once in the group, he does not need a reputation and no one needs to know about his relationship to other members. All they need to know is whether he has any money left. I said that group altruism requires rather complex behavior, but actually it is not all that complex—as long as a group of organisms can count, they don't have to watch one another or coordinate closely.

All of the rules of group altruism that I have described in the previous section are followed by a money system. To check this, let us imagine some of the organisms from the previous group as people who are using money. Organism N+1N+1 can be Alice and organism N+2N+2 can be Bob. The problem is to show how the group uses money to encourage Alice to coordinate with the rest to punish and reward altruism in Bob.

First, suppose Bob deserves to be rewarded. This means that he has acted altruistically toward some group member. He has done a favor and received money for it. Now the group requires Alice to reward Bob by behaving altruistically toward him. If she does, and she performs a favor for him in return for money, then she gets a reward for this—she can use the money to redeem a favor from someone else. On the other hand, if she does not reward Bob, then she is not rewarded either. She doesn't get the money and therefore cannot redeem favors. Now suppose, that Bob failed to be altruistic to the group, and needs to be punished. When it is Alice's turn, all she has to do is refuse to serve him. Her reward for doing so is simply the opportunity cost of performing a favor without being rewarded. On the other hand, if she is altruistic toward Bob, then her punishment is that she receives no money from him, and hence cannot redeem a favor later from someone else.

Note that the “punishment” in this style of game is nothing more than an opportunity cost. Physical punishments and social pressure are not necessary. People who do not coordinate with the group simply forgo future altruism from group members. In fact, Alice may very well have reasons to behave altruistically toward Bob though he has no money. They might develop their own system of reciprocal altruism between the two of them instead. If the money system was not useful enough, then everyone would revert to that, which would apparently be a kind of gift economy.

Earlier in this article I described a paradox of money. How is it possible for everyone to want money because everyone wants money because everyone wants money, etc.? Once most of a population begins to behave this way, then anyone who does not is punished by being excluded from the money economy. We all must value money, even though it may not appear to make any rational sense and the whole concept may seem fraught with paradox. This does not explain how monetary behavior begins, but it explains why it persists once it is established, and it does so by reference to the individual benefit of the behavior. This answer may seem trivial or obvious, but it is not: how many people go around thinking of the money in their pockets as a means of punishing one another for not being altruistic enough? We all know that it is horrible not to have any money, but this theory gives an explanation as to why.

This is not to say that charity is bad or that paupers deserve to be miserable. To show that something is functional is not the same as justifying it absolutely. Rather, there is a reason that a money system works and why it should be expected to remain successful. Radically different arrangements are likely to fail against it, and social theory must take this into account.

I have argued in an earlier section that monetary behavior is a form of altruism and I have shown here that it satisfies one of the conditions of reciprocal altruism—that of punishing non-altruists, or of preventing them from benefiting from the altruism of everyone else. The other condition required of reciprocal altruism is that there should be lots of expected future opportunities to call in favors. In monetary economics, the word for this is liquidity, and it roughly refers to the demand for money. A money is more liquid the more easily a given sum of goods can be sold for money without markedly altering its demand. If a money is liquid, therefore, then many opportunities for redeeming favors all around. If money is illiquid, then society might run out of favors to give before one runs out of money, which means that the money must become less valuable. Thus, as with other forms of altruism, therefore, money has a network effect. If lots of people want lots of money, then the system is more effective than when few people want it.

What Is Good Money?

If money is a form of reciprocal altruism, we can have a very good idea of what makes good money. Good money is whatever is closest to an ideal system of reciprocal altruism. It is easy to see how many of the traditional qualities of good money serve to enable this ideal. I’ll just list a few here. Some of these qualities have been attributed to Aristotle, but this is a misattribution, and I don’t know where they originally came from.

A money system just needs to be an association of a number with each person that they can transfer to one another, and which can be used to track a person’s altruism. Money must be scarce because if it were not, then anyone could easily increase the size of their number associated with them without having to provide favors for them. Money must be difficult to counterfeit because if it were not, it that just means that it is hard to tell what the correct number associated with a person should be. Money must be fungible, divisible, and durable because then it wouldn’t act like a number at all. Finally, money must be portable because then it could not be easily transferred and would necessarily be less liquid. In a global economy, in which people do business from opposite ends of the world, money should not only be portable, but teleportable.

The earliest form of money was commodity money, which means that the monetary unit was a good whose scarcity is a result of the physical difficulty of producing it. Before the days of mass production, money could be produced

from materials that were not scarce or valuable because the scarcity and inefficiency of human labor kept them in short supply. For example, wampum was a kind of hand-made bead produced from clam shells. Other forms of commodity money, such as gold, are scarce because there is a strictly limited amount on Earth.

Commodity money has the advantage of allowing for cooperation between people who have no knowledge of one another. It can enable trade routes between Europe and China during times when the Europeans have no idea what China is like and vice versa. It is the only form of money possible when communication between different peoples is limited. Even knowing nothing of the technology or culture of different lands, probably no one there can cheaply manufacture gold and silver. This assumption has sometimes backfired; during the age of exploration, European settlers and merchants caused monetary crises in the lands they visited because the natives relied on money that they could not mass produce, but which the Europeans could. This was the fate of wampum and the famous Yap stones.

The problems with commodity money is that it is not very portable and costly to protect from theft. People have often resorted to storing their commodity money in secure warehouses, which came to be called banks. If a bank is well-known, people can trade in receipts which grant them the right to redeem the commodity money from the bank. This leads us to a second form of money.

Fiat money is a kind of money that exists because it is stamped with the brand of an institution. An institution issues units of currency, and it can make as much as it likes. The units may be imprinted on physical pieces of paper or may be nothing more than numbers in a ledger. It is the responsibility of the institution to ensure that it continues to function properly as money. The integrity of fiat money is backed by the reputation of the issuing institution.

Historically, fiat money has arisen out of banknotes whose convertibility into commodity money was gradually reduced to nothing. If this is done slowly enough, people continue to use the bank notes as money even though they are no longer backed by a commodity. This is a dishonest practice, but fiat money is not necessarily dishonest. We could imagine an honest fiat money that was never issued under false pretenses, though this is certainly not the historical norm. The result is a kind of money which is much more portable than commodity money but whose scarcity is protected by the promise of the issuing institution rather than the physical difficulty of producing it. Fiat money requires a greater degree of communication between a people for it to work. Someone is certainly not going to accept pound notes if he doesn't know what England is or dollars if he has never heard of the USA.

The principle of concentrated benefits and distributed costs has meant that issuers have often been able to take advantage of users by manipulating the money supply to their benefit. The issuers of money can effectively evade punishment for a failure to be sufficiently altruistic by creating extra units of money at a very low cost. Unlike regular folk, who must never run out of money, the issuers can continue to spend and never run out.

Finally, money as a distributed system is the latest step in the evolution of money[x]. This is what Bitcoin is. The principle which protects Bitcoin from manipulation is the difficulty of changing the behavior of the distributed system. Money as a distributed system has no physical presence at all. It is one huge ledger, one which must be duplicated over many locations. It must be impossible for someone to change a copy without being detected, and there must be some process in place to ensure that all copies remain in agreement.

Money as a distributed system requires a much greater degree of communication between everyone in order to function properly. Fiat money requires that there be one well-known institution, but a money as a distributed system requires everyone to be in communication with everyone else. If this is possible, its advantages are many. The supply can be strictly limited—not just to some unknown amount of something in the Earth, or to an amount that an issuer can change, but to a specific number that no one can change. Trusting a well-designed distributed system is more like trusting the laws of physics than trusting a human agent. Because it can be protected by cryptography, it can be made very difficult to steal at a low cost. Therefore I would conclude that a distributed system is capable of coming closest to the ideal of perfect delayed reciprocal altruism.

Monetary Fallacies

Reciprocal altruism is a great first start as a theory of money because it so neatly undercuts a lot of the most common fallacies. First, what gives money value? An adherent of commodity money might say that it is the industrial uses of the money good, whereas an adherent of fiat money might say that it is the force of the government issuing it, and the loyalty people have toward their government. Neither of these answers is true. It is true that some system is required to keep track of who has money and who does not, but that is not what makes money valuable. The value of money is the value of cooperation. It is that simple. The value of money is not somehow in the monetary unit; it is in the whole of society and in peoples' desire to cooperate.

Second, is there value in changing the supply of money? Evidently not; changes to the money supply only prevent the money from functioning as a form of reciprocal altruism. Whoever first received the new money would be able to spend it without having done anything to earn it. Such a money system allows non-altruists to benefit. It allows for cheating. It would be possible, of course, for all money everywhere to multiply by the same proportion at the same time. This could be done for real in a distributed system. If that happened, no one would benefit disproportionately. But what would be the point? Doing so clearly changes nothing about the real state of the economy and would have only the effect of immediately changing all prices by the inverse proportion.

Finally, is there anything inherently antisocial about being wealthy or about amassing wealth? If making money means just granting favors, then it is not inherently antisocial no matter how much money someone makes. The only question is whether it is possible to give out favors that are not as beneficial as they may at first appear, but this is a very subjective question that would arise in any system of cooperation. A person who has a high income is just someone able to provide very useful favors. A person with a large savings is even better; he is someone willing to donate to the economy without demanding much back. If he spends his money eventually, then he was giving to the economy on a loan; if he is willing to maintain his savings permanently, then he is willing to remain permanently invested in the future productivity of the economy. This is social behavior, not antisocial!

In an economy like ours with billions of humans all cooperating with one another, some people will figure out ways of doing favors for billions of people all at the same time and consequently will end up with thousands or millions of times the amount of money that people have on average. Even this cannot be criticized. The possibility of becoming one of those extra-helpful individuals is part of the benefit of altruism, and it is therefore part of the incentive that fosters cooperation. If it is curtailed with a system of progressive taxation, then that reduces the everyone's willingness to cooperate, not just those paying the tax.

This is not to say that there may not be other reasons for progressive taxation, but not because excessive wealth is somehow damaging to the social order. Rather, the contrary is true: progressive taxation will tend to make people less cooperative, at least as far as the money system is concerned.

There are, however, other systems of cooperation that humans rely on. All of them require investment by the participants, and they are therefore all in competition with one another. I think that the extraordinary benefits of money explain much about why group ideologies like religions, political parties, and nationalist ideologies maintain teachings which are so averse to it. Both the money economy and ideological groups are enablers of group altruism. Consequently, they are in competition with one another. Both require investment of their members' time and effort, and a person can only give so much. Yet the money economy gives such a superior return for most things than do ideological groups that it leaves the groups at a severe disadvantage. When you hold a large cash balance, you are investing in the whole human race rather than in one group, and you are invested in a system capable of a vastly greater degree of specialization and invention. Such steep competition would threaten the cohesion of any ideological group.

Conclusion

That monetary behavior is a form of altruism, and of reciprocal altruism in particular, has been demonstrated by

relating the properties of money to the definition of reciprocal altruism and to its required features. The theory of reciprocal altruism is not, however, a complete theory of money. It leaves some things out. In the first section of this article, I explained that importing a concept from biology into economics requires that it be treated according to a very different theory of value. I have not done so here. That is what is missing from a complete theory of money. In particular, a subjective theory of value allows for two new phenomena not treated here.

First, the services which most animals can perform for one another are very limited. Studies of reciprocal altruism in the wild usually look at only two goods: a single kind of favor which the animal give one another, and the debt that animals owe to one another for having received it. In a human economy, there are many, many goods. Money tends to result in a set of unitary prices for all goods, but the theory of reciprocal altruism does not explain price formation and why there should be a unitary set of prices.

Second, the value of money can change relative to the entire economy. In other animals, altruism evolves as an instinct over generations, whereas a human economy can rapidly change in the degree to which it rewards cooperation. A human can get the idea that his money is going to be worth more or less relatively soon and change his behavior accordingly in anticipation. Money cannot be fully understood without developing these topics.

Special thanks to Jonathan Vaage of BTC Design for the diagrams.

Footnotes

i. Dawkins, R., *The Selfish Gene*, Oxford University Press, 1976. ↩

ii. Szabo, N., "Shelling Out: The Origins of Money", The Satoshi Nakamoto Institute, 2002, SHA256:30f7eea01d4b60c3ca33b6337a32b391. ↩

iii. See Smith, J., *Evolution and the Theory of Games*, Cambridge University Press, 1982, SHA256:88fdfb61c32b2758863e7f338b97a7e9 for the classic exposition of this idea. ↩

iv. The truth is actually more complicated—in biology, individuals and groups exist on many levels. Within a cooperative colony of animals, we would need to explain why an individual animal cannot benefit at the expense of everyone else; but each animal is itself a eusocial colony of individual cells which are extremely regimented in how they cooperate with one another. So there is also a need to explain why all the cells are better off cooperating with one another to be an animal that cooperates with others of its own kind instead of all turning into individualistic cancers. And a cell, too, functions because the many genes within it all cooperate very nicely instead of competing to replicate out of control.

   See Dawkins and Trivers, R., *Social Evolution*, The Benjamin/Cummings Publishing Company, Inc., 1985 for decisive criticisms of group selection. The idea of group selection has recently been revived under the name multilevel selection theory. See (put citation here) I'm not a professional biologist, so take this for what it's worth, but I think that this article rather misses the point. ↩

v. Trivers, R., "The Evolution of Reciprocal Altruism" , vol. 46, no. 1, Quarterly Review of Biology, 1971, pp. 35-57, SHA256:4c2dfb4acc63492b06151bce8db7c452. ↩

vi. Stevens, C., "Modeling Reciprocal Altruism", vol. 47, no. 4, British Journal for the Philosophy of Science, 1996, pp. 533-551, SHA256:7847130bd62ba99c21b3a83fdf528662. ↩

vii. For a delightful introduction to game theory, see Luce, R., Raiffa, H., *Games and Decisions: Introduction and Critical Survey*, Dover Publications, 1989. For an introduction that is also free, see Watson, J., *Strategy: An Introduction to Game Theory*, 3rd ed., W. W. Norton & Company, 2013, SHA256:2ce0ccbea55ca88c908b876f22bbb379. ↩

viii. Stevens, 1996, pp. 533-551. ↩

ix. Trivers, 1985. ↩

x. I follow Graf, K., "Bitcoin Decrypted Part III: Social theory aspects" , KonradSGraf.com, 27 Dec 2013 in treating cryptocurrencies as a separate category from commodity money or fiat money. ↩