Altruistic Behavior: Lessons from Neuroeconomics

Kei Yoshida
Postdoctoral Research Fellow
University of Tokyo Center for Philosophy (UTCP)

Table of Contents

1. The Emergence of Neuroeconomics, or the Decline of Homo Economicus

The Idea of Homo Economicus, or Economic Man

1. Fundamental assumption in modern economics
2. Coined by J. S. Mill's critics (J. K. Ingram and J. N. Keynes)
3. Basic ideas
    1. An agent is rational, self-interested, and emotionless.
    2. The agent works for maximizing his/her utility.
4. Typical Criticism: based on a mistaken psychological view

A Defense of Homo Economicus

Milton Friedman, "The methodology of positive economics" (1953)

1. As far as theories can provide accurate predictions of economic behavior, unrealistic assumptions are not problematic (*a kind of pragmatism).

2. We cannot test theories by examining whether their assumptions are realistic.

Mainstream economists do not abandon the idea of homo economicus. But there are a lot of problems with homo economicus. Game theoretic situations are good examples.

The Ultimatum Game

1. Player A is asked to split 20 monetary units with player B.
2. A makes an offer to B (for instance, A=15; B=5)
    1. If B accepts, then both receive the proposed money respectively.
    2. If B declines, then both receive nothing.
3. Game theoretic prediction is as follows:
    1. B will accept any offer because he/she is rational and selfish.
    2. A also anticipates this, and thus offers 1 monetary unit to B.
4. But the case is different from the game theoretic prediction!

A Discrepancy between Theory and Reality

"Obviously, a selfish homo oeconomicus will offer his opponent the smallest share possible because game theory suggests that the responder will accept it. Yet in experiments conducted by game theories *the most frequent outcome is a fair (i.e. 50:50) share*. Moreover, several studies report that *in about 50% of all games responders who are offered some 20% off the total amount choose to reject the offer* although this means missing out altogether." (Kenning and Plassmann 2005, 348; Italics added)

Why?

Typical explanation: A's offer was not fair, and thus B was offended. Because of that, B rejected the offer. When we say that B was offended, we refer to B's emotion. What does it mean? Here comes neuroeconomics as a new field of inquiry.

2. Neuroeconomic Studies of Altruistic Behavior

Rilling and Others on Social Cooperation

Rilling, J. K., D. A. Gutman, T. R. Zeh, G. Pagnoni, G. S. Berns, and C. D. Kitts. 2002. A neural basis for social cooperation. *Neuron* 35: 395-405.

Used the iterated prisoner's dilemma game and scanned 36 female subjects with fMRI

Two players A and B are asked whether they would cooperate with each other, and each is offered some amount of money that is based on the chosen interaction of the round.

Rilling and Others on Social Cooperation

There are four possibilities:

Both players cooperate (CC); player A cooperates, but player B defects (CD); player A defects, but player B cooperates (DC); both players defect (DD); Hence the payoff matrix is this.

		Player A	
		Cooperation	Defect
Player B ( )	Cooperation	2 (2)	3 (0)
	Defect	0 (3)	1 (1)

Rilling and Others on Social Cooperation

Rilling and others did two experiments.

In the 1st experiment, they scanned 19 subjects during each of four sessions and tried to "isolate the neural correlates of cooperation and noncooperation in social and nonsocial contexts, and of monetary reinforcement of behavior." "The results of the first experiment revealed different patterns of neural activation depending on whether the playing partner was identified as a human or a computer" (ibid.).

Thus Rilling and others scanned 17 subjects to investigate differences between human and computer interactions during each of three sessions.

Rilling and Others on Social Cooperation

The results of the experiments:

The subjects tend to cooperate, and "there was a tendency for subject pairs who arrived at a CC outcome to persist with mutual cooperation so that a CC outcome in the current round was most likely to be followed by a CC outcome in the next" (396).

"[T]he largest activation for this interaction involving symmetric social behavior is in the anteroventral striatum and sub-general anterior cingulate cortex (BA 25). The striatal activation includes the caudate nucleus accumbens (Nac), *both of which receive midbrain dopamine projections known to be involved with processing reward*. The ventromedial/ orbitofrontal cortex (OFC), *another brain area involved in reward processing*, was also activated for the interaction" (397; Italics added).

Rilling and Others on Social Cooperation

The activated brain areas are related to processing reward, and thus the subjects choose to cooperate feel pleasure.

Rilling and others compared mutual cooperation with a human partner and that with a computer.

The result: "[i]n neither of the two experiments did mutual cooperation with a computer activate the rostral anterior cingulate or the anteroventral striatum observed for human playing partners" (398).

Hence these two areas seem to be related to cooperative interactions with human partners, and the subjects derive pleasure from such interactions (a kind of sociality?)

Sanfey and Others on Economic Decision-Making

Sanfey, A. G., J. K. Rilling. J. A. Aronson, L. E. Nystrom, and J. D. Cohen. 2003. The neural basis of economic decision-making in the ultimatum game. *Science* 300: 1755-8.

Scanned 19 subjects who played the ultimatum game as responders, and investigated how their brains worked when they had fair and unfair offers

According to Sanfey and others, bilateral anterior insula, dorsolateral prefrontal cortex (DLPFC), and anterior cingulate cortex (ACC) were activated for unfair offers from human partners than for fair offers from human partners. Furthermore, the activation for unfair offers from human partners was greater than that for unfair offers from computer partners.

Sanfey and Others on Economic Decision-Making

This result seems to suggest that social contexts are relevant to the activations of the brain areas because the subjects are more sensitive to unfair offers from human partners.

The activation of bilateral anterior insula is interesting in that it is associated with negative emotions such as anger, disgust, distress, and pain. Moreover, “those participants with stronger anterior insula activation to unfair offers rejected a higher proportion of these offers” (1756-7; Fig. 3 [A]).

By contrast, DLPFC is relevant to cognitive processes such as accumulating money. “Unfair offers that are subsequently rejected have greater anterior insula than DLPFC activation, whereas accepted offers exhibit greater DLPFC than anterior insula” (1757; Fig. 3[B]).

Sanfey and Others on Economic Decision-Making

Furthermore, ACC that is linked to cognitive conflict was activated in the cases of unfair offers.

From all of these things, Sanfey and others suggest that there may be a conflict between emotions (anterior insula) and cognition (DLPFC) when the subjects have unfair offers, and ACC deals with such a conflict.

Important: both anterior insula and DLPFC were activated in the cases of unfair offers!

de Quervain and Others on Altruistic Punishment

de Quervain, D. J.-F., U. Fischbacher, V. Treyer, M. Schellhammer, U. Schnyder, A. Buck, and E. Fehr. 2004. The neural basis of altruistic punishment. *Science* 305: 1254-8.

Investigated whether there is any neural basis of altruistic punishment of defectors in the trust game.

Using PET, scanned 14 subjects when they played the trust game

de Quervain and Others on Altruistic Punishment

What is the trust game?

1. Players A and B are given 10 monetary units respectively.
2. A makes a decision of whether he/she would keep 10 MUs, or trust B and send 10 MUs to B.
    1. In the former case, both receive 10 MUs.
    2. In the latter case, the experimenter quadruples the money (i.e. 40) and gives 40 MUs to B. Thus B has 50 MUs; however, A has nothing.
3. B is asked whether he/she would give A back half of 50 MUs (i.e. 25) or keep all of the money.

de Quervain and Others on Altruistic Punishment

4. If B keeps all the money, B betrays A. A takes it as a norm violation.

5. A is offered the option of punishing B by assigning up to 20 punishment points to player B, and A has one minute to decide whether he/she punishes B and how much punishment points should be used. de Quervain and other scanned the subjects during the one minute period.

6. There are three ways to punish the defectors.

    1. For 1 punishment point, one MU for A and two MUs for B (costly)

    2. For 1 punishment point, nothing for A and two MUs for B (free)

    3. For 1 punishment point, nothing for A and B (symbolic)

de Quervain and Others on Altruistic Punishment

de Quervain and others compared effective punishment with symbolic punishment. Effective punishment reduce the defectors' economic payoff; however, symbolic punishment does not.

In effective punishment, the caudate nucleus (the dorsal striatum) that is associated with rewarding process was activated.

This suggests that the punisher derive satisfaction from punishing the defector. Furthermore, those who with stronger caudate activation tend to spend more money to punish the defectors.

3. What Can We Learn from Neuroeconomics?

Ernst Fehr's View of Strong Reciprocity

Ernst Fehr: leading behavioral/neuro-economist; works on human altruism

"Strong reciprocity is a combination of altruistic rewarding, which is a predisposition to reward others for cooperative, norm-abiding behaviours, and altruistic punishment, which is a propensity to impose sanctions on others for norm violations. Strong reciprocators bear the cost of rewarding or punishing even if they gain no individual economic benefit whatsoever from their acts. . . . Strong reciprocity thus constitutes a powerful incentive for cooperation even in non-repeated interactions and when reputation gains are absent, because strong reciprocators will reward those who cooperate and punish those who defect." (Fehr and Fischbacher 2003, 785).

This differs from what we call altruism. By altruism, we usually mean the intention of increasing the welfare or well-being of others.

A Biological View of Altruism

Fehr and his collaborators adopt a biological view of altruism, not a psychological view of it.

"According to the biological definition, an act is altruistic if it is costly for the actor and confers benefits on other individuals. It is completely irrelevant for this definition whether the act is motivated by the desire to confer benefits on others, *because altruism is solely defined in terms of the consequences of behavior*. This contrasts with the psychological definition, which also requires that the act be driven by an altruistic motive that is not based on hedonic rewards" (de Quervain et al. 2004, 1257; Italics added).

Altruism is defined only in terms of the consequences of behavior. Hence an agent's intention is not a matter.

A Psychological Criticism of Fehr

Mark Peacock's criticism of Fehr (Peacock 2007)

1. According to Fehr's explanation, altruistic punishment is motivated by negative emotions such as anger.

2. By ascribing non-material subjective satisfaction to a punisher, Fehr returns to the selfish axiom that he criticizes.

We can find these points in Sanfey and others (2003) and de Quervain and others (2004). Thus Peacock's criticism is prima facie plausible.

Is Altruism Really Altruistic?

Then is altruistic punishment or altruism in general really altruistic? Is it merely hypocritical or even vicious?

These are difficult questions to answer; however, there are two things to be mentioned.

1. Peacock fails to discern that neuroeconomics criticize the idea of homo economicus: a rational, self-interested, and emotionless agent (cf. Sanfey and others)

2. Even if altruistic behavior turns on emotions, this would not necessarily undermine the very idea of altruism. Is altruistic behavior without emotions really possible? Probably, no.

The Problem with Strong Reciprocity

We have thus far seen how we can meet Peacock's criticism of Fehr. But strong reciprocity does not guarantee that human beings are always altruistic.

A small number of selfish individuals suffice to make altruists defect. "If strong reciprocators believe that no one else cooperate, they will also not cooperate" (Fehr and Fischbacher 2003, 787).

Hence it is important to organize a social institution in such a way that prevents selfish individuals from behaving selfishly. Punishment is one way. "[E]ven a minority of strong reciprocators suffices to discipline a majority of selfish individuals when direct punishment is possible" (ibid.).

One Lesson from Neuroeconomics

In my view, this is one of the important lessons from neuroeconomics. Even if we are strong reciprocators, we do not always behave altruistically. Whether we behave altruistically turns on our social institutions. Hence we need to organize a social institution in such a way that promotes altruistic behavior and punishes selfish behavior.

In this sense, we are still responsible for our behavior.

The End