

Using the Coefficient of Rationality to Allow Classical Game Theory
To Handle Emotions, Cultural Norms and Non-Perfectly Rational Players
by Mark Stuckel

Abstract: Current advancements in game theory, namely Rabin's incorporation of fairness, have substantially improved game theory's applicability to real world interactions by recognizing that humans do not act perfectly rationally. However, it will be demonstrated that these improvements are incomplete. By understanding that *all* human emotions affect how rationally we act (positively *or* negatively), we can better grasp real-world strategic interactions and better formulate the effects various emotions have on them. Once this is recognized, it will become possible to measure to what degree emotions detract or enhance a person's rationality. After every emotion and cultural norm is accounted for, we will be left with a final numerical value called the Coefficient of Rationality (CoR). The CoR is defined as the degree to which a person's actions conform to those of a perfectly rational being. The CoR should not be seen as a gauge for normative analysis, instead it should be seen as a *predictive* tool which incorporates emotions and cultural norms into GT in order to determine Nash Equilibriums for non-perfectly rational players.

1. Introduction

The overall purpose of this paper will be to introduce the concept of the Coefficient of Rationality and to explain how it will allow us to use classical game theory to generate corresponding Nash Equilibria for players who act from emotions and cultural norms and who are not perfectly rational. Additionally, several philosophical questions will be addressed along the way.

I will employ the following strategy:

- 1) Explain what classical game theory says about the theoretical outcome of the Ultimatum Game

- 2) Show what *actually* happens in real world Ultimatum Games.
- 3) Discuss how Rabin improved upon GT by incorporating fairness.
- 4) Show why Rabin's model is incomplete.
- 5) Provide a better model for incorporating emotions and other cultural norms into game theory by introducing the concept of the Coefficient of Rationality.
- 6) Show how the CoR is measured and used to find adjusted Nash Equilibria.

I: What does classical game theory say about the theoretical outcome of the Ultimatum Game?

The Ultimatum Game is an experimental game used in economics to model strategic interaction. The first player (player A) is awarded a sum of money and must divide it by making an offer to the other player (player B). Player B can either accept the offer or reject it. If B accepts the offer, the money is split as dictated by A. If B rejects the offer, neither player gets any money.

In classical game theory, meaning that if the game were played by two perfectly rational beings desiring only to maximize their own monetary payoff, the resulting outcome would be player A will offer 1 cent to player B and 99 cents to himself. The reason being that since player B has no say in the division, his best option is to just accept whatever positive offer is given to him. If he rejected the 1 cent he would not be maximizing his utility since he would then be forgoing a positive gain. The Nash Equilibrium is therefore A, 99 : B, 1. A Nash Equilibrium occurs when both players are currently playing a strategy that cannot be improved upon by switching to a different strategy. We will later see that the CoR will help determine adjusted Nash Equilibria for non-perfectly rational beings that incorporate emotions, cultural norms and other preferences.

II: What do the results from experiments by Kahneman, Knetsch and Thaler say about the *actual* outcome of the Ultimatum Game?

The conclusion reached by Kahneman, Knetsch and Thaler is that people tend to offer fairly even splits and when an unfair split is offered (usually less than a 25% offer) the offer is rejected.

TABLE 2 Experiment 1 Results¹

Class:	Psychology/ Psychology	Psychology/ Commerce	Commerce/ Psychology
Mean amount offered (\$)	4.76	4.47	4.21
Equal split offers (%)	81	78	63
Mean of minimum acceptable (\$)	2.59	2.24	2.00
Demands > \$1.50 (%)	58	59	51
Participants (<i>IV</i>)	43	37	35

As a result of their research, Kahneman, Knetsch and Thaler offered an even broader conclusion. They stated: “The traditional assumption that fairness is irrelevant to economic analysis is questioned. Even profit maximizing firms will have an incentive to act in a manner that is perceived as fair if the individuals with whom they deal are willing to resist unfair transactions and punish unfair firms as some cost to themselves. The rules of fairness...help explain some anomalous market phenomena.”¹

Their paper marked an important step in the improvement of game theory. By recognizing that humans do not interact in accordance with classical game theory, he set the stage for Rabin to mathematically formalize fairness and incorporate it GT in order to better model real world strategic interactions. The crucial step was to realize that most people take more than just utility and monetary payoff into consideration when making decisions in the Ultimatum Game (and life in general). For example, some people are strongly influenced by what others will think about their actions and therefore act accordingly. Others are influenced by their deep religious beliefs or lack thereof. It will later be shown that these emotional and

cultural inclinations affect how “rational” a person is (rational meaning maximizing their own payoff). And by formalizing these inclinations mathematically, it will be possible to complete Rabin’s model by incorporating all emotions and cultural norms into GT. When this is complete, it will be possible to use classical game theory to develop adjusted Nash Equilibria based off of non-perfectly rational players. But before this can be discussed, we have to first understand how Rabin formalized just one aspect: fairness.

III: How did Rabin formalize the concept of fairness and thus improve the real-world applicability of classical game theory?

Firstly, Rabin reformulated Kahneman, Knetsch and Thaler’s view by saying, “People like to help those who are helping them, and to hurt those who are hurting them. Outcomes reflecting such motivations are called fairness equilibria. Outcomes are mutual-max when each person maximizes the other’s material payoffs and mutual-min when each person minimizes the other's payoffs. It is shown that every mutual-max or mutual-min Nash equilibrium is a fairness equilibrium. If payoffs are small, fairness equilibria are roughly the set of mutual-max and mutual-min outcomes; if payoffs are large, fairness equilibria are roughly the set of Nash equilibria.” He specifically addresses the Ultimatum Game in his paper on page 1284 (The American Economic Review, December 1993). One of his main points was to show that in real world Ultimatum Games, classical game theory strategy did not maximize payoffs.

To summarize briefly for clarity:

A *Nash equilibrium* occurs when both players are currently playing a strategy that cannot be improved upon by switching to a different strategy.^{4, 5}

Mutual max outcomes occur when, “given the other person's behavior, each person maximizes the other's material payoffs.”² [p. 1282]

Mutual min outcomes occur when, “given the other person's behavior, each person minimizes the other's material payoffs.”²[p. 1282]

Rabin goes beyond simply recognizing that fairness considerations affect people’s behavior and instead puts forth a method/framework to formalize fairness mathematically to incorporate it into game theory. Rabin’s purpose was to take this new formalization and apply it to the payoff matrix to show that people alter their behavior based on how they are being treated. For example, people who believe they are being treated unfairly will treat the aggressor badly in return. Conversely, people who feel they are being treated well will return the favor. This fluctuation in the way people interact is described by fairness equilibria.

With this conceptual schema in mind, he bases his framework off of a similar idea developed by Geanakoplos, Pearce and Stacchetti in order to formalize fairness. The framework developed by GPS took standard game theory and modified it to incorporate actions *and* beliefs when calculating payoffs.

Rabin dismisses the argument that this new formulation is unnecessary and that these beliefs could somehow be incorporated by “transforming payoffs” and analyzing it in the conventional way. Rabin’s response is to say that if one were to try and formalize this under classical game theory, it would inevitably lead to various contradictions. For example, in the “battle of the sexes” game, the standard game theory model says that each player “strictly prefers to play his strategy given the equilibrium.”² This means that there needs to be a way to formalize the fluctuations of one person's payoffs based on:

- 1: The husband’s strategy.
- 2: The wife’s beliefs about what her husband’s strategy will be
- 3: The husband’s belief about what his wife believes his strategy will be.

The following formula incorporates these beliefs about fairness into GT to better understand the desired Nash Equilibrium of the husband and wife:

$$U_i(a_i, b_j, c_i) = \pi_i(a_i, b_j) + F_j(b_j, c_i) \cdot [1 + f_i(a_i, b_j)] \quad ^2(\text{page 1287, Vol. 83 No. 5})$$

So in plain English we can read each of these pieces as follows:

U_i = Player I's utility

a_i = the husband's or player i's strategy

b_j = The wife's or player J's belief about what player I's strategy is going to be

c_i = player i's belief about what the other player believes his (player i) strategy is

$\pi_i(a_i, b_j)$ = player i's payoffs

$F_j(b_j, c_i)$ = how kind player i thinks player j is being to him

$f_i(a_i, b_j)$ = how kind player i is being to j

IV: Why is incorporating fairness, on its own, into game theory an incomplete approach?

The short and simple answer would be to say that although Rabin's fairness formulation is a significant step forward in the understanding of real world interactions by incorporating a psychological element into game theory, he fails to see that fairness is only one human emotion or cultural trait that affects how people interact. For example:

- 1: Other human tendencies beyond fairness affect the outcomes of the Ultimatum Game.
- 2: There can be varying degrees of fairness (e.g. variations across cultures).
- 3: Some people do not act fairly (see table 2).

V: What is a better way to incorporate fairness and other human tendencies into GT?

The key step in improving upon Rabin is to recognize that fairness is something that is not (perfectly) rational. By rational, it is meant that one maximizes utility or pay off or expected outcome (I do not mean rational in a larger ethical or philosophical sense). With the perfectly

rational being as a reference point, we notice that emotions affect people's actions either positively or negatively. Instead of looking at fairness as something by itself, we should instead look at it as something that adds or subtracts a certain measurable quantity X from the perfectly rational level of 1.

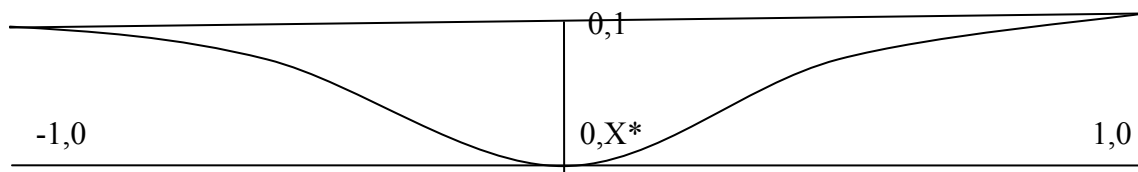
These levels of rationality are measured by the Coefficient of Rationality.

1: The CoR is a numerical measure of how rational an entity is on a scale from:

a: +1 (perfectly rational; acts as a perfectly rational being would)

b: 0 (perfectly random)

c: -1 (perfectly irrational; does exact opposite of perfectly rational being)



The x-axis stands for the coefficient of rationality.

The y-axis stands for the probability we can predict the agents actions/beliefs. (note*: the lower bound probability, where the curve crosses the y axis, is defined as 1/the number of choices. In the Ultimatum Game, that probability for player A is 1/101 since there are 101 amounts to offer. The probability for B in the Ultimatum Game is 1 / 2 since he has to choose either accept or reject. The upper bound is 1, hence the straight line across the top.)

We can read the chart and apply it to the Ultimatum Game as follows:

1: A, 1 : B, 1 both agents are perfectly rational, CoRs of 1.

A,99 : B,1 the outcome will be as dictated by standard game theory. It is important to note, that the perfectly rational agent always maximizes their outcome/payoffs regardless.

2: A, 1 : B,-1 (A acts perfectly rationally, B acts perfectly irrational.)

A, 100 : B, 0 (A will know that the most irrational thing to do as player B would be to accept an offer 0.)

3: A, -1 : B, 1

A, 0 : B, 100 (A being perfectly irrational would offer the lowest expected outcome which would be 0 and B being perfectly rational would gladly accept.

4: A, -1 : B, -1

A, proposes 0 : B, 0 (he rejects the 100 payoff. Notice that when both players are perfectly irrational, the result is the lowest possible outcome: neither player gets anything)

5: A, 0 : B, 0

A, 25 : B, 25 (note: for the perfectly random ones, I took the limit of the payoffs as if the game were played an infinite amount of times. For example in this case, A will offer an average of 50 and B will accept $\frac{1}{2}$ of the time, thus leading to an overall pay off of 25 for both.

6: A, 1 : B, 0

A, 50 : B, 0 (Knowing that B is perfectly random, he knows that B will accept exactly $\frac{1}{2}$ the time regardless of what A offers. Therefore A offers 100 every time. He has nothing to gain to by lowering his bid, and nothing to lose by always offering 100.

7: A, -1 : B, 0

A, 0 : B, 50 (The perfectly irrational player will always offer 0, and player B will accept the 100 half of the time.

8: A, 0 : B, 1

A, 49 : B, 49 (Player will offer an average of 50 and player B will accept every offer except for an offer of 100. Therefore, each player loses that one round.

9: A, 0 : B, -1

A, 1.01 : B, 0 (Player will offer an average of 50 and player B will reject every offer except for an offer of 0. Therefore, each player A wins that one round and get's 100. Winning 100 out of every 101 turns equals an overall payoff of 1.01.

Now that the extremes are covered, we can work backwards and fill in the gaps for the other values, thus leading to some important conclusions:

- 1: The more rational a person is, the more likely they will accept a low offer, and the more likely they will offer a higher number.
- 2: Acting irrationally has a lower payoff outcome than acting randomly.
- 3: The higher the sum of the CoRs, the greater the expected outcome for both players.
- 4: The lower the sum of the CoRs, the lower the expected outcome for both players.
- 5: The only times the full prize is handed out every round, is when at least one of the players is perfectly rational.
- 6: The only time no money is distributed every round, is when both players are perfectly irrational.

With these numbers in mind, we can now view fairness as something that, all things being equal, subtracts a certain value from a CoR of 1. The beauty of this approach compared to Rabin's is that now we can add in other human tendencies into the mix and determine how they affect someone's CoR. Depending on whether they affect it positively or negatively, the new CoR value will affect how the Nash Equilibrium is adjusted.

My fundamental claim is that all of our emotions and cultural norms will have a certain measurable effect on the rationality of our actions. Since this section might be controversial, I will first address some of the philosophical implications of this approach.

In the extant and voluminous literature dealing with human emotions, there is much debate over how certain traits affect a subject's actions. In the previous paragraph, I have claimed that, all things being equal, fairness would negatively affect a person's expected outcome in the Ultimatum Game. However some have argued that fairness and other emotions can also benefit a person's expected outcome (like Jack Hirschleifer's *On the Emotions as Guarantors of Threats and Promises*.) In his article, Hirschleifer argues that people can

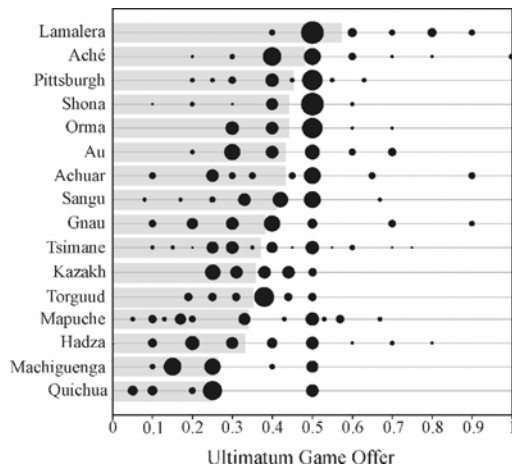
sometimes improve their expected outcome by not acting self interestedly. Additionally, "... emotions can serve a constructive role as guarantors of threats or promises in social interactions" [OEGTP, p.2]. Nevertheless, this fact does not take away from the view I am advancing.

In the previous paragraph, I stated that all things being equal, fairness would negatively affect a person's expected outcome in the Ultimatum Game. However, if we include "all other things," it could very well be the case, that fairness can improve a non-perfectly rational being's expected outcome.

One of the philosophical implications the CoR will have for fields dealing with "all other things"(emotions, cultural norms, etc.) is that it will become clear that the correct way to view the effects of emotions and other cultural norms will be to view them in terms of how they will affect a person's CoR. Doing so will determine which emotions positively or negatively affect a person's utility. In addition to this, it will also determine which emotions affect utility the most or least in different situations. For example, fairness might equal, on average -.1 for all situations. While equaling -.2 in the Ultimatum Game and +.1 in a true bargaining situation at a street market. Generosity might be on average -.1 and selfishness might be on average +.1. But what we can say with absolute certainty from table 2 is that the students are not acting perfectly rationally. It then follows that their emotions and other cultural norms as a whole are affecting their decisions and outcomes and subsequently their CoRs. If they were acting perfectly rationally, the experimental results would perfectly reflect classical game theory. Once the list of the traits and their effects is large enough, it would be possible to add them all together and come up with average values for different populations.

For example, several research studies³ have been conducted to show how different cultures play the Ultimatum Game differently. From the chart below, it can be observed that

people in different societies with different social and cultural norms will offer different amounts to player B. Similarly, (if we assume player B is perfectly rational) the farther down the list we go, the higher the populations CoR i.e. the closer they conform to what a perfectly rational person would offer.



“A bubble plot showing the distribution of UG offers for each group. The size of the bubble at each location along each row represents the proportion of the sample that made a particular offer. The right edge of the lightly shaded horizontal gray bar gives the mean offer for that group. Looking across the Machiguenga row, for example, the mode is 0.15, the secondary mode is 0.25, and the mean is 0.26.”³

As a side note, before there is any confusion, I would like to make clear that the CoR should be seen as a purely predictive tool to better find Nash Equilibria and not as a scale with which to compare and judge right from wrong, virtuous from non-virtuous or good emotions from bad emotions. The CoR will be used to better assess how different people with different norms and emotions interact with one another. It is beyond the scope of this paper to determine whether or not game theory should or could be used for normative analysis in fields like ethics or value theory. The CoR should not be seen as the gold standard of how one *should* act, but rather as a guide for maximizing desired payoffs. And, as we will see shortly, this includes both monetary and non-monetary payoffs.

VI: How do we measure the CoR? And how, specifically, is it used?

By conducting tests similar to those in table 2 we can determine that the average person strays from perfectly rational behavior by X%. If the results showed that average amount awarded was 90% of the total prize pool, we would say they strayed from perfectly rational behavior by 10% and had an average CoR of .8.

$$\text{CoR} = 1 - (\% \text{ difference to perfectly rational outcome} * 2)$$

$$\text{CoR} = 1 - (.10 * 2)$$

$$\text{CoR} = .8$$

We subtract from 1 because that is the value assigned to a perfectly rational group. We multiply the % by two in order to extend the spread into the negative numbers for the perfectly irrational side.

Another way to determine CoRs would be to use economics experiments specifically designed to determine how close a give person compares to what a perfectly rational person would do in the given situation. For example, you could hand them a survey that asks questions along the lines of: Would you accept or reject an offer of 50? How about 40? 30?...1? Also, what would be the highest value you would offer to a person who is perfectly rational? Etc, etc... And after collecting enough data you could develop a bell curve that would graph the CoRs and determine a mean, median and standard deviation. Then you would be able to find a desired Nash Equilibrium that corresponds to player B's desires which also maximizes player A's desires, too.

One possible objection to this might be to arguing the following point. How does the CoR complete Rabin's model of incorporating fairness into the problem of measuring and formalizing non-utility based payoffs? Rabin seems to be saying that for most people, they do

not act 100% rationally because they “feel better” when they act fairly. Thus fairness is a non-utility payoff that must be factored into game theory when trying to determine and maximize ones monetary and non monetary payoff. The CoR model only focuses on how much the two players end up making in the end and not their overall non-utility based payoff associated with forgoing additional monetary gain in return for a sense of wellbeing by acting fairly.

Now, this is a legitimate objection but I argue that it is misguided. The CoR does take into considerations the player’s preferences for fairness and other non-utility based payoffs. When the previously mentioned survey is taken, it can be assumed that a player who refuses to offer less than 40%, prefers this action because their emotions or cultural norms tell him that being fair, or whatever he wants to call it, is “worth” more than the forgone monetary gain.

One might object to this again on the grounds that the CoR couldn’t possibly differentiate between a person who is very altruistic and a person who simply just isn’t very rational. Both people would be assigned the same CoR. The objector would continue, “But shouldn’t we be concerned with *why* people are acting a certain way?”

In response to this, I would that yes, the CoR cannot differentiate between these two cases. However, since GT is only concerned with the ends and payoffs, the “means” of how we achieve these ends does not matter. For example, what difference does it make that a person will not offer a bid lower than 40 because he fears person B will reject his offer and he will lose out on the money vs. a person who will not offer less than 40 based off of a strong moral conscience and a desire to always be fair? The purpose of the CoR is to mathematically formulate all emotions and cultural norms into classical game theory in order to find Nash Equilibria that are adjusted to incorporate emotions, cultural norms, and non-utility based desires. Therefore the

reasons for people's actions aren't as important as the outcomes of their actions and with this, the problem of incorporating non-utility payoffs is solved.

Rabin formulated fairness into game theory to better model how people interact with one another and how fairness changed the Nash Equilibria. One of his findings was that people tend to want to help others who are helping them and hurt others who are hurting them. His formalization of fairness introduced a model that fluctuated to different equilibriums depending on how player A was treating player B.

Now that we have the CoR value, it raises the questions, "Ok, now what do we do with it? Why is it important?" And in response, there are several answers. The most important ones include:

- 1) By applying this value to two individuals, we can better formulate a corresponding Nash Equilibrium based on their CoRs.
- 2) The more information we know about a particular person or population, the better we can predict how they will respond/act.
- 3) This value is extremely important because if we could find a value that predicts with great accuracy how a population will interact with itself, then we could find the best "offers" to make to maximize our desired payoffs. Additionally, by applying this framework beyond the Ultimatum Game into the business world, or other social environments, we could better model various strategic and not strategic interactions.

To illustrate this last point I will now show through an example how to use the CoR to find an adjusted NE that maximizes each player's desired payoff. Let us first imagine the following possible data for a typical Ultimatum Game.

	X	Nash Equilibrium	X
Offer =>	50/50	30/70	10/90

Accepts offer	B 50, A 50	B 30, A 70	B 10, A 90
	99% chance = 49.5avg	75% = 52.5avg	20% = 18 avg
Rejects offer	0,0	0,0	0,0

*** Let's assume that a person with a .7 CoR will accept a 50/50 split 99% of the time, and a 30/70 split 75% of the time and a 10/90 only 20%. The CoR is what helps us determine how likely player B will accept or reject a particular offer. ***

If we put ourselves in person A's position, the person making the offer, we can quickly find the offer that maximizes our expected outcome and the offer that represents a Nash Equilibrium.

First we calculate our own CoR (since we have the most information about ourselves). For this example I will use a CoR of 1 for the player making the offer. Next, if we know that the average person to which we are offering a portion of the prize has a CoR of .7, we can optimize our utility/payoff by offering a value that maximizes what we give to ourselves and minimizes the rejection rate. Therefore, from this chart we can infer that optimal offer is 30, because it will return to us an average of 52.5. Knowing player B's CoR helped me to determine his acceptance rate. I cannot improve my payoff by changing my offer and player B doesn't gain by rejecting. Therefore, we are at an adjusted Nash Equilibrium that could not have been found using classical game theory alone. The CoR was instrumental in allowing us to calculate the NE for non-perfectly rational players by converting their emotional and cultural inclinations into a numerical value. This value was then used to determine their acceptance/rejection rate, and then supply player with the knowledge of how to maximize his payoff.

VII: Application beyond the Ultimatum Game

A better understanding of the CoR would lead to a more unified model of how different people with different emotions, cultural norms and preferences interact with each other. It is not hard to imagine the possible effects this could have on businesses, advertisers and other fields

involved in the maximization of payoffs as a result of strategic interaction. Along these lines, studies of this nature could change how politicians “market” themselves to the electorate. This paper should provide a solid foundation upon which future game theorists, mathematicians and psychologists could base a coherent formulation of the human psyche, in order to better predict strategic interactions among individuals and groups, thus maximizing a player’s desired outcome. Additionally, this paper should introduce new philosophical questions and implications. For example, should the CoR be used to make normative statements about which emotions or preferences are or are not rational?

VIII: Conclusion

The classical model of game theory is insufficient for describing real world strategic interactions primarily because it is designed to handle perfectly rational beings; humans, however, are not perfectly rational. Rabin improved upon this by mathematically formalizing fairness into GT. However his formalization was incomplete because he failed to incorporate other influential emotions and cultural norms. By expanding upon Rabin’s fairness model using the CoR, it will now be possible to mathematically incorporate these emotions and cultural norms into GT in order to determine Nash Equilibriums for non-perfectly rational players.

Bibliography

- 1) *Fairness and the Assumptions of Economics*, by Daniel Kahneman, Jack L Knetsch and Richard H Thaler, *The Journal of Business*, Vol. 59, No. 4, Part 2: The Behavioral Foundations of Economic Theory. (Oct., 1986), pp. S285-S300.
- 2) *Incorporating Fairness into Game Theory and Economics*, by Matthew Rabin, *The American Economic Review*, Vol. 83, No. 5. (Dec., 1993), pp. 1281-1302.
- 3) *“Economic man” in cross-cultural perspective: Behavioral experiments in 15 small-scale societies*, Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, Richard McElreath, Michael Alvard, Abigail Barr, Jean Ensminger, Kim Hill, Francisco Gil-White, Michael Gurven, Frank Marlowe, John Q. Patton, Natalie Smith, David Tracer. 2005.
- 4) <http://www.gametheory.net>, various articles, accessed February 20 – April 21, 2008
- 5) “Game Theory,” <http://plato.stanford.edu/entries/game-theory/>, accessed February 20 – April 21, 2008
- 6) “Rationality and Utility from the Standpoint of Evolutionary Biology” by Donald T. Campbell. *The Journal of Business*, Vol. 59, No. 4, Part 2: The Behavioral Foundations of Economic Theory. (Oct., 1986), pp. S355-S364.
- 7) “Comments on the Interpretation of Game Theory” by Ariel Rubinstein. *Econometrica*, Vol. 59, No. 4 (July, 1991), 909-924