SOME PROBLEMS WITH JUDGING RATIONALITY[1]

Piotr Swistak[2]
University of Maryland

Abstract: *The gap between game-theoretic predictions and actual choices people make in, for instance, gaming experiments has been over-interpreted as evidence against rationality of players. I consider a version of the ultimatum game and examine its equilibria under different assumptions about players' preferences. Using standard notions of rationality I show that the discrepancy between the "normative" and the "descriptive" cannot be established by a simple comparison of what is predicted by the equilibrium choices and the actual choices people make.*

Key words: *game theory, behavioral game theory, ultimatum game, rationality.*

PEWNE PROBLEMY Z OCENĄ RACJONALNOŚCI

Streszczenie: *Różnica między predykcjami wynikającymi z teorii gier a rzeczywistymi wyborami dokonywanymi przez uczestników eksperymentów jest nadinterpretowana jako dowód na nieracjonalność graczy. W pracy rozważam pewną wersję gry ultimatum i badam jej równowagi przy różnych założeniach o preferencjach uczestników. Korzystając ze standardowego pojęcia racjonalności wykazuję, że rozbieżność między wynikiem „normatywnym" i „deskryptywnym" nie może być ustalana poprzez proste porównanie punktów równowagi gry z rzeczywistymi wyborami dokonywanymi przez ludzi.*

Słowa kluczowe: *teoria gier, behawioralna teoria gier, gra ultimatum, racjonalność.*

 An earlier version of this paper was presented at the Public Choice Society Meetings, San Antonio, Texas, March 12-15, 2015. The author thanks an anonymous reviewer whose comments were most helpful in cleaning a number of slips in an earlier version of the paper.

[2] Piotr Swistak, Department of Government and Politics and Applied Mathematics, Statistics, and Scientific Computation Program, University of Maryland, College Park, MD.

 DOI: 10.7206/DEC.1733-0092.44

1. Introduction

It will be fair to say, I believe, that the comparison of actual choices with game theoretic predictions – both in casual observations and in the more systematic findings of behavioral game theory – are contained in one or more of the following statements: "people are not rational," "people are not self-interested," "real choices are (very) different from game theoretic predictions." Statements like that have been perpetuated by game theorists, both from experimental and theoretical circles, and widely and frequently repeated by conventional social scientists who typically have a very bleak understanding, if any at all, of what rationality means or may mean.

While we cannot do much to ameliorate the problem of conventional social scientists using undefined concepts, we should be very careful not to slip into the same trap ourselves. If those who know the proper definitions of terms used in behavioral game theory do not clearly communicate the meaning of these terms to others then we cannot blame the others for doing the same to the public at large.

In this paper I will argue that claims regarding the meaning of empirical findings in gaming experiments should be qualified in a more careful way than they usually are. I will consider a series of games, their equilibria and some related data and use it as an excuse to reflect on the notion of rationality, self-interest, and what is often interpreted as a discrepancy between what people choose and what game theory predicts they should choose. My objective is to make us aware of some interpretations of this discrepancy that are perhaps non-obvious yet quite consequential for a proper understanding of what we see in human behavior in general and behavioral game theory in particular.

Below I will occasionally refer to observations I made in my game theory classes. I should emphasize that my data were not collected in any regular fashion nor were they obtained through properly designed experiments. They should, hence, be treated as anecdotal evidence at best. Given the nature of my data I will not use them as evidence of anything but rather as a pretext to talk about proper and improper interpretations of similar results that have been observed in studies in behavioral game theory. The running example that I will use is a specific version of the ultimatum game.

2. The ultimatum game – an informal description

One exercise I have used in all my classes, literally, is the following version of the ultimatum game:

 DOI: 10.7206/DEC.1733-0092.44

Suppose a wealthy benefactor, of the Sergey Brin variety, wants to donate 100 million dollars to two universities he has attended – the University of Maryland and Stanford. Sergey prefers that the division of the money is settled by the interested parties rather than himself. All he is willing to do is to set the rules under which the two parties will negotiate the division. Sergey arranges a three-way teleconference with the presidents of the two universities – I will call them Mary and Stan (tacky choices are easier to remember) – and explains how and what the two will be allowed to negotiate.

To simplify the process of negotiations each side will have only one shot at proposing a division and one shot at accepting or rejecting the other's proposal. All divisions are to be comprised of whole millions only. More specifically, Sergey asks Mary and Stan to proceed as follows. Mary will go first and propose a division of 100 million to Stan. Stan can either accept or reject the proposed division. If Stan accepts Mary's proposal, the allocation becomes final and binding for all three parties. If Mary's proposal is rejected, Sergey reduces the stake to 90 million (think of 10 million reduction being a penalty for not reaching an agreement or a cost of Sergey's time spent on prolonged haggling) and Stan gets his turn at proposing a division of 90 million to Mary. If Mary accepts the division, the allocation is made as proposed. If Mary rejects Stan's proposal, Sergey withdraws his offer altogether and both universities end up with nothing.

Suppose that Mary and Stan have just heard about the intended donation, and the negotiation rules, during a live three-way teleconference with Sergey. As the live feed continues Sergey expects them to promptly carry out the negotiation. As specified by the rules Mary is supposed to make an initial offer. Can you say what Mary will propose? Will Stan accept Mary's offer or will he turn it down and go on to split a smaller pie proposing his own division? How much money will they end up with?

This version of the ultimatum game is more complex than ultimatum games used in empirical studies. In the pioneering study of Güth et al. (Güth, 1982), and in dozens of studies that followed (cf. Camerer C.F., 2011; Camerer and Thaler, 1995) one player proposes a division and the other player either accepts or rejects it and in the case of rejection they both get noting. Such "one stage negotiation" design is very useful since it does not burden subjects with any difficulty involved in solving for equilibrium and, hence, provides a clean and simple measure of their social preferences. Since my objectives are different, the ultimatum game I consider

does not follow the same design. One of the issues I am interested in, for instance, is whether students can identify a solution under some well-defined, or nearly-well defined, set of assumptions regardless of how they would behave in such a negotiation themselves. To put it differently, I am interested if students can identify a reasoning of a rational, self-interested person once we define what a rational, self-interested player is.

So, how should a rational or a self-interested individual behave in the ultimatum game?

3. The Ultimatum Game – A Proper Specification

The negotiation puzzle, just like nearly all social science problems, requires that we carefully identify all the assumptions under which we seek the solution to the problem. If we don't do that then asking people to solve it will conflate their ability to identify the assumptions with their ability to find a proper solution under these assumptions. In fact, in some classes I have asked students to solve the problem before I have explicitly stated the assumptions while in other classes, after the assumptions were stated. Interestingly, the answers were quite similar. A minority of students predicted a lopsided split of (10, 90), (11, 89), (1, 89) or (9, 91) while a majority predicted a more egalitarian split with a much smaller difference between the two payments. Divisions like (40, 60), (50, 50), (60, 40) and others where the difference between the two payments is not greater that 20 would be typical of those predictions.

One reason why it makes sense for me to ask for prediction before setting up the proper specification of the problem is that the details of the story I tell are supposed to make the necessary assumptions overt and obvious.

For instance, since Mary and Stan are university presidents they should both try to raise as much money as they possibly can regardless of how much money will be left for the other institution. In fact, this is what is required by their job description. If Stan, for instance, makes a mistake and brings back home less than he could have, someone on the board of directors is likely to point it out to him and perhaps ask for his resignation. For Mary and Stan, self-interest is not an option – it is the necessity of their job. Yet, no matter how obvious the self-interest assumption may seem, we should begin by specifying its content.

Self-interest of Mary and Stan concerns their preferences over the set of possible outcomes of the negotiation. The outcome of the negotiation is a pair of payments, for instance, (20, 80), (80, 10), (0, 0), of which – as the notation I use here will assume –

 DOI: 10.7206/DEC.1733-0092.44

the first payment in a vector (M, S) goes to Mary and the second to Stan. The set of all outcomes includes all possible integer divisions of 100 million: (100, 0), (99, 1), (98, 2), …, (2, 98), (1, 99), (0, 100), all possible integer divisions of 90 million: (90, 0), (89, 1), (88, 2), …, (2, 88), (1, 89), (0, 90) and, finally, (0, 0). With the set of possible outcomes identified, the assumption we made about Stan and Mary can be stated as follows:

> A1. (SELF-INTEREST/EGOISM) If Mary's payment in outcome (M, S) is larger than her payment in outcome (M*, S*), i.e., M > M*, Mary will prefer the first outcome over the second; an analogous assumption applies to Stan.

The next assumption concerns the rules of negotiation. These rules are stated in a clear and unambiguous way by Sergey and if we wanted to define them in a formally proper way we would have to specify an extensive form game that begins with a decision node of Mary with 101 branches taking off from it, each corresponding to a different proposed division of 100, etc, and ending with terminal nodes in which players get specific payoffs.

PAYMENTS VERSUS PAYOFFS It is important to note that in game theory the concept of a payoff denotes player's utility of an outcome – not the payment received in the outcomes. For this reason the distinction between the "payment" and the "payoff" is very important. For instance, it may well be the case that Mary prefers outcome (90, 10) to outcome (90, 0). If she does, then we say that her utility of the first outcome is larger than her utility of the second one or, equivalently, that the first payoff is larger than the second. Of course, the payments in both, outcomes, (90, 10) and (90, 0), are identical. This shows that payments and payoffs are distinct notions.

And so the extensive form game consists of the game tree and players' payoffs in terminal nodes. The game tree itself can be thought of as a model of the negotiation rules. Since we have not specified the payoffs or the assumptions that would imply them, we have to leave them out of the picture for a moment and restrict our attention the game tree only. The game tree alone, as is suggested by the next axiom, corresponds to the rules of the negotiation process.

> A2. (THE RULES OF NEGOTIATION) Extensive form game tree corresponding to the negotiation rules specified by Sergey.

The next assumption concerns what players actually know about the game. As was the case with A1 and A2, this assumption is clearly embedded in the negotiation story: Given that all information is given by Sergey during a three-way teleconference, all of it, as would be reasonable to assume, becomes common knowledge. Common knowledge, which itself is a rather difficult to define concept in game theory can be intuitively described as follows: An information is a common knowledge in a group

of players if they all know it, and they all know that they all know it, and they all know that they all know that they all know it, and so on, ad infinitum. Given the teleconference setup, the rules of negotiation, i.e., the content of A2, are clearly common knowledge. The fact that self-interest is common knowledge is implicit in the assumption that Mary and Stan are university presidents and their jobs require them to act in the best interest of their institutions. Thus, we can now state the third assumption behind the ultimatum game puzzle.

> A3. (COMMON KNOWLEDGE) All assumptions made about the game are common knowledge.

At this point we may be tempted to start solving the game. This, however, may be premature. A closer look at the first assumption will make us realize that it does not specify players' preferences over the entire set of outcomes. For instance, A1 tells nothing about Mary's preference between (10, 90) and (10, 80). In general, we don't know players' preferences over two outcomes in which they get the same amount of money while the other player does not. This seemingly tangential consideration is very important. If we do not know player's preferences over the entire domain of alternatives, we don't know if his *preference relation is rational* (a strict preference relation is *rational* if it is asymmetric and negatively transitive[3]) or not. In consequence, we cannot specify player's utilities of the different outcomes. In game theory these utilities are called payoffs and if we don't have payoffs we don't have a game. What is more, with incomplete preferences not only we don't know what the payoffs are – we don't even know if they exist! And so, the problem of the missing assumption is neither negligible nor marginal. In a sense, as we will see next, it will turn out to be important in our considerations.

Assumptions A1, A2 and A3 are all a straightforward formalization of conditions overtly imbedded in the story defining the ultimatum game. The additional assumption we need to make in order for the preference relation to be properly defined will not be in this category at all. If we were to conjecture what Stan will do when choosing between (10, 90) and (0, 90), there is no uniquely obvious assumption that comes to mind. If Stan were malevolent, for instance, then he would prefer (0, 90) over (10, 90) since in the second outcome Mary is left with less money. If Stan were

[3] Preference relation $\prec$ is *asymmetric* on D if for any a, b $\in$ D, if a $\prec$ b then $\neg$(b $\prec$ a). Relation $\prec$ is *negatively transitive* on D if for any a, b, c $\in$ D if $\neg$(a $\prec$ b) and $\neg$(b $\prec$ c) then $\neg$(a $\prec$ c). Symbol $\neg$ denotes negation. Preference relation satisfying these two conditions is sometimes referred to as a rational preference relation or simply preference relation. This concept of rationality can be equivalently defined with a weak preference relation $\preceq$ ("prefers or is indifferent to") and assuming that $\preceq$ is *connected*: for any a, b $\in$ D either a $\preceq$ b or b $\preceq$ a, and *transitive*: for any a, b, c Î D if a $\preceq$ b and b $\preceq$ c then a $\preceq$ c.

 DOI: 10.7206/DEC.1733-0092.44

benevolent he would prefer to leave Mary with more and would hence pick (10, 90) over (0, 90). If Stan were completely unconcerned with Mary's welfare he would be indifferent between (0, 90) and (10, 90) and may choose to pick each with probability 0.5. There may be some other reasonable assumptions one may want to consider but I propose that we stick to the three suggested above.

Let's begin with specifying the malevolence assumption. For the reasons that will soon become clear I will not label this assumption as A4 but rather as A4.1.

> A4.1. (MALEVOLENCE) If Mary's payments in outcomes (M, S) and (M*, S*) are the same, i.e., M = M*, Mary will prefer the outcome in which Stan gets less; an analogous assumption applies to Stan.

THE CONCEPT OF A PAYOFF AND THE NOTION OF RATIONALITY With A4.1 in place the set of outcomes is well ordered by each of the player's strict preference relation. In other words, A1 and A4.1 imply that the preference relation is asymmetric and negatively transitive. A preference relation that is asymmetric and negatively transitive is called rational. This notion of rationality is equivalent to the existence of a utility function which is, informally speaking, an assignment of numbers that reflects players' preferences. More specifically, Mary's utility of any outcome (M, S) is a number $u_M(M, S)$ that satisfies the following condition: for all outcomes (M, S) and (M*, S*), $u_M(M, S) > u_M(M^*, S^*)$ if and only if Mary prefers (M, S) over (M*, S*). It can be proved that such utility function exists if and only if Mary's preferences are asymmetric and negatively transitive. In game theory these utilities are called *payoffs* and they constitute a defining element of a game. Without payoffs we don't have a game.

Asymmetry and negative transitivity imply the existence of the so called ordinal payoffs which merely denote the order of alternatives. For example, a set of payoffs 0, 1 and 2 is formally the same as the set -0.1, 2.17 and 103. To put it differently, ordinal payoffs are invariant under any strictly increasing transformation.

And so, with assumption A4.1 the specification of the game is complete. We thus have the first full set of assumptions, a few more sets to come, and a better idea on what we may and may not be able to infer from an observation that, say, 30% of subjects predicted (10, 90), (11, 89), (1, 89) or (9, 91) while 70% predicted some outcome with the payment difference of no more than 20 like, for instance, in (40, 60), (45, 55) or (60, 40).

 DOI: 10.7206/DEC.1733-0092.44

4. Predictions – assuming rationality and self-interest

Once we have a properly specified problem, prediction is nothing more than a deduction made from the set of assumptions that define the problem. What we need to do next, then, is to consider what follows from the four assumptions we have just made:

A1. (SELF-INTEREST/EGOISM) If Mary's payment in outcome (M, S) is larger than her payment in outcome (M*, S*), M > M*, Mary will prefer the first outcome over the second; an analogous assumption applies to Stan.

A2. (THE RULES OF NEGOTIATION) Extensive form game tree corresponding to the negotiation rules laid out by Sergey.

A3. (COMMON KNOWLEDGE) All assumptions made about the game are common knowledge.

A4.1. (MALEVOLENCE) If Mary's payment in outcomes (M, S) and (M*, S*) are the same, M = M*, Mary will prefer the outcome in which Stan gets less; an analogous assumption applies to Stan.

Note that A3 assumes that A1, A2 and A4.1 are all common knowledge. From this set of assumptions we can now derive a prediction.

GAME 1: Under assumptions A1, A2, A3 and A4.1 the prediction is (10, 90).

The proof of this claim goes as follows: If the game goes to the second step, Stan will maximize the amount he gets by offering 1 million to Mary and keeping 89 for himself. To see why consider the following two points: First, if Stan offers to give 0 to Mary and keep 90 for himself, Mary will reject the proposal because rejection results in both getting 0 and given that Mary ends up with 0 in both cases but Stan gets less in the second case, by A2, Mary will prefer outcome (0, 0) to (0, 90). Second, by A1, Mary will accept the proposed division of (1, 89) since the alternative, (0, 0), gives her less. Now, since Mary knows that if the game goes to the second step she will end up with 1 million and Stan with 89, in step one of the game she cannot offer Stan less than 89. An offer of (11, 89) will maximize what Mary can get if Stan accepts it but if he rejects it, the game will go to step two and, as we already know, in step two Stan will end up with 89 anyhow. Thus, we should ask: would Stan take 89 in step one and leave Mary with 11, or would he go to step two, and leave her with 1? Since Stan is malevolent (as assumed by A2) he would prefer that Mary gets 1 rather than 11. But, by A3, Mary can make the same inference that we have just made which would make her offer in the first step not (11, 89) but (10, 90). This offer will maximize her share of the division. By A1, Stan will accept the (10, 90) proposal.

 DOI: 10.7206/DEC.1733-0092.44

Before I comment on the prediction it may be interesting to ask what would happen in the game had we replaced the malevolence assumption A4.1 with an opposite assumption of benevolence (A4.2.)

> A4.2. (BENEVOLENCE) If Mary's payment in outcomes (M, S) and (M*, S*) are the same, M = M*, Mary will prefer the outcome in which Stan gets more; an analogous assumption applies to Stan.

Just like the assumption of malevolence, A1 and A4.2 imply that players' strict preference relations are rational in a sense of being asymmetric and negatively transitive.

Having explained in detail the prediction under the first set of assumptions, I often ask my students whether they think the prediction will change if instead of malevolence we assume benevolence. More than 90% say that prediction will change. But, will it? As it turns out, it will not.

GAME 2: Under assumptions A1, A2, A3 and A4.2 the prediction is (10, 90).

Even though the prediction, as we will see, remains unchanged the reasoning, of course, must be different, since the set of assumptions is different. These assumptions imply different payoffs which means that we have here a different game. Under this set of assumptions the reasoning will look as follows: If the game goes to the second step, Stan will maximize the amount he gets by offering 0 to Mary and keeping 90 for himself. Mary will accept Stan's proposal because if she rejects it then both will get 0 and given that she ends up with 0 in both cases but Stan gets more in the first, by the benevolence assumption A4.2, Mary will prefer (0, 90) over (0, 0). Now, since Mary knows that if the game goes to the second step she will end up with 0 and Stan will end up with 90, in step one she cannot offer Stan anything less than 90. Clearly, (10, 90) will maximize what she can get, but we need to consider if Stan is going to accept it. If Stan were to turn down (10, 90), then the game would go to the second step and, as we already know, Stan would get 90. Thus, we should ask: Would Stan take 90 in step one and leave Mary with 10, or in step two and leave Mary with 0? Since from A4.2 Stan is benevolent he would prefer that Mary gets 10 rather than 0. Knowing this Mary will offer (10, 90) in the first step and Stan will accept her offer.

As I have mentioned before, the content of A4, whatever we assume it to be, cannot be derived from the description of the game. This makes A4 stand out from other assumptions as arbitrary and artificial. For this reason it would be prudent to look at what happens under different versions of A4. One assumption we should certainly consider is that players' preferences are independent of the payments to the other player or, equivalently, that players are indifferent between outcomes that give them the same payment. Hence, for Mary, for instance, $u_M(M, S) = u_M(M, S^*)$ for all S and S*

 DOI: 10.7206/DEC.1733-0092.44

which means that Mary's utility function is only a function of the payment to her i.e., $u_M(M, S) = u_M(M)$ for all M and S. A4.3 which specifies this assumption will be the last form of A4 we will look at. In the absence of other information, it will be reasonable to assume that when choosing between two outcomes that have equal utilities a player is equally likely to choose any one of them. Hence we adopt the following axiom A4.3.

> A4.3. (INDIFFERENCE THAT IMPLIES EQUAL PROBABILITIES OF CHOICE) If Mary's payments in outcomes (M, S) and (M*, S*) are the same, i.e., M = M*, Mary is equally likely to choose either outcome; an analogous assumption applies to Stan.

If we want to make a prediction under A4.3 we will have to address the following consideration. If the game goes to the second step, and Stan offers (0, 90), by assumption A4.3 Mary will be equally likely to reject the offer, which would end the game with (0, 0), and accept it, which would end it with (0, 90). For Stan this means that he will get 0 with probability ½ and 90 with probability ½. So, should he go for this option or should he rather offer Mary (1, 89) in which case he would be sure that Mary will take the offer? In other words Stan has to choose between 89 for sure and a lottery in which he gets 0 with probability ½ and 90 with ½. Which one should he opt for?

With the question just asked the issue of rationality changes substantially. Recall that defining rational choice under certainty, which was all we needed in Games 1 and 2, only required that we have an asymmetric and negatively transitive, i.e., rational, preference relation. (Existence of a rational preference relation is equivalent to the existence of ordinal payoffs.) This notion of rationality, however, applies to decision making under certainty. When choices involve risk, it is, obviously, insufficient.

The standard notion of rationality in decision making under risk assumes a player with asymmetric and negatively transitive preference relation that also satisfies axioms of continuity and independence. In other words, *rationality in decision making under risk* is defined by the axioms of the Von Neumann and Morgenstern's expected utility theory.[4] Adding the axiom of continuity and independence is very consequential in that it changes the nature of payoffs in the game. When only the preference theory was assumed payoffs were merely ordinal indices of preferences. With the new axioms added this is no longer the case. Now payoffs 0, 1 and 2 are no longer equivalent to payoffs 1, 2 and 4, for instance. (They were identical under the preference theory since they designated the same order.) Under expected utility theory payoffs are no

[4] Denote by [a, p; b, 1-p] a lottery in which alternative a is obtained with probability p and alternative b with probability 1-p. Expected utility theory assumes the following three axioms about a strict preference relation $\prec$: $\prec$ is asymmetric and negatively transitive; *independence*: for any a, b, c ∈ D and any p ∈ (0, 1] if a $\prec$ b then [a, p; c, 1-p] $\prec$ [b, p; c, 1-p]; and *continuity*: for any a, b, c ∈ D if a $\prec$ b $\prec$ c then there exist p, q ∈ (0, 1) such that [a, p; c, 1-p] $\prec$ b $\prec$ [a, q; c, 1-q].

 DOI: 10.7206/DEC.1733-0092.44

longer invariant under any strictly increasing transformation, they are only invariant – as has been proved by Von Neumann and Morgenstern – under strictly increasing linear transformations. This means that numbers representing payoffs not only reflect the order of preferences but they also reflect their intensity. And so, as we move from the preference theory to the expected utility theory payoffs change their nature from ordinal scale to interval. Hence under the expected utility theory payoffs acquire a profoundly different meaning. For this reason in game theory we make a distinction between *ordinal payoffs* (preference theory) and *cardinal payoffs* (expected utility theory.) These two types of utilities correspond to two different concepts of rationality that are used in game theory and in decision making.

Since A1 thru A4.3 do not state anything that will allow us to conclude that players are rational in the sense of expected utility theory if we want to make predictions under A1-A4.3 we have to add the following assumption to this set.

> A5. (EXPECTED UTILITY THEORY) Players' preference relations are asymmetric and negatively transitive and satisfy axioms of continuity and independence. In short, players' preferences satisfy axioms of the Von Neumann and Morgenstern's expected utility theory.

Keeping in mind that we deal with different type of payoffs we will return now to making the prediction. So, should Stan choose 89 for sure or a lottery in which he gets 0 with probability ½ and 90 with ½? To answer this question we have to turn to Stan's utilities for 90, 89 and 0 or, more specifically, we have to know if his utility of 89, $u_s(89)$, is larger or smaller than $½\ u_S(0) + ½\ u_S(90)$. But this depends on Stan's utility function. If, for instance, $u_S(x) = 2^x$ then $u_S(89) < ½\ u_S(0) + ½\ u_S(90)$ but if $u_S(x) = \sqrt{x}$ then $u_S(89) > ½\ u_S(0) + ½\ u_S(90)$. Assumption A1 means that players' utility functions are increasing in money but it does not say anything beyond that – and neither do axioms of the expected utility theory (A5). Are players' utility functions concave, convex, linear? Again, to be able to move forward with our prediction in this game we have to conjecture something about the utility functions. One common assumption in micro economics is that of a concave utility function. Let's suppose, then, that both players have concave utility functions and assume that in addition to A1 and A2, A4, A5 and A6 also become common knowledge among players.

> A6. (CONCAVE UTILITY FUNCTIONS) Players' utility functions are concave.

With concave utility function Stan will prefer 89 over a lottery in which he gets 0 with probability ½ and 90 with ½ which means that in the second step Stan will offer (1, 89) and Mary will take the offer. Now, since Mary knows that if the game

 DOI: 10.7206/DEC.1733-0092.44

goes to the second step she will end up with 1 and Stan will end up with 89, in step one she cannot offer Stan anything less than 89. Will, however, Mary's offer of (11, 89) work? Since Stan is indifferent between (11, 89) and (1, 89) he is equally likely to leave Mary with 11 and with 1 which leaves her with the dilemma of getting 10 for sure if she offers (10, 90) or getting 11 with probability ½ and 1 with ½. Given that Mary's utility function is concave, she will prefer 10 over the lottery. This, however, means that she will offer (10, 90) and Stan will accept the offer. And so, we can finally formulate the prediction.

GAME 3: Under assumptions A1, A2, A3, A4.3, A5 and A6 the prediction is (10, 90).

We may still consider what will happen if Mary's and Stan's utility functions were convex, say $u(x) = 2^x$, and extend this direction of model specification further[5] but instead I propose that we consider what happens if we vary A4 in yet a different way and assume that Mary and Stan have different preference relations. Since solutions in the next two cases use the same reasoning that has already been used three times before, I will omit the explanations and limit the presentation to new versions of assumption 4 and the predictions that follow.

> A4.4. (MALEVOLENT MARY AND BENEVOLENT STAN) If Mary's payment in outcomes (M, S) and (M*, S*) are the same, i.e., M = M*, Mary will prefer the outcome in which Stan gets less. If Stan's payments in outcomes (M, S) and (M*, S*) are the same, i.e., S = S*, Stan will prefer the outcome in which Mary gets more.

GAME 4: Under assumptions A1, A2, A3 and A4.4 the prediction is (11, 89).

> A4.5. (BENEVOLENT MARY AND MALEVOLENT STAN) If Mary's payments in outcomes (M, S) and (M*, S*) are the same, i.e., M = M*, Mary will prefer the outcome in which Stan gets more. If Stan's payments in outcomes (M, S) and (M*, S*) are the same, i.e., S = S*, Stan will prefer the outcome in which Mary gets less.

GAME 5: Under assumptions A1, A2, A3 and A4.5 the prediction is (9, 91).

For a moment I will leave these predictions without a comment. I will return to them in the concluding section. In the next step I would like to entertain one other possibility in which we leave the assumptions of rational choice (preference theory and the expected utility theory) intact but do not insist that players are self-interested.

[5] Changing this assumption is, in fact, going to change the prediction. I leave it to the reader to see why the prediction will be affected and how.

 DOI: 10.7206/DEC.1733-0092.44

5. Predictions – assuming rationality but relaxing self-interest

One easy, almost automatic way to relax the self-interest assumption is by adopting the opposite assumption of pure altruism. This can be done by reversing the condition of self-interest and correcting it for the case of equal payments. More specifically,

> A1*. (ALTRUISM) If Stan's payment in outcome (M, S) is larger than his payment in (M*, S*), i.e., S > S*, Mary will prefer the first outcome over the second regardless of how much she gets. If Stan gets the same amount in both outcomes, i.e., S = S*, Mary will prefer the one in which she gets more. An analogous assumption applies to Stan.

GAME 6: Under assumptions A1, A2 and A3 the prediction is (90, 10).*

To derive the solution to Game 6 it is enough to use the reasoning we have used before. I will leave this exercise to the reader.

Non-selfish behavior seems to be at the core of empirical findings not just with ultimatum games (cf., Camerer C. F., 2011; Davis, 1993.) Our version of altruism as defined by A1* is the most extreme version of non-selfishness. But there are other, more balanced, ways of modeling non-selfish behavior. One of them I will explore in more detail.

When I ask students to tell me why they chose a "close-to-egalitarian" division I often hear that it is important to them what the other person gets or that fairness of the division is important to them[6] or that if they offer the other player too little he is bound to reject such offer[7]. When, in reply, I ask if they would agree to represent such consideration by putting some weight on their own payment and some (i.e., the remaining) on the other player's payment, they readily agree that such model would properly represent their thinking. While I do not want to suggest that there is any credibility to such admissions, I do find it interesting to check deductive consequences that follow from the suggested model and see how they compare with actual choices. This is what I will do next.

One straightforward way to model a "weighted payoff" consideration is by assuming that Mary's utility of a division (M, S) is wM + $(1-w)$S where w is Mary's degree of egoism and 1-w is her degree of altruism and assume the same about Stan.

[6] As Kaminski et all (2013) note: "Subjects in […] experiments can clearly see that their actions reveal something about their norms and values and this means that they may not have sufficient motivation to reveal their true attitude. In fact, there is strong evidence showing that in studies in which subjects are guaranteed anonymity their "concern for fairness" drops dramatically (E. K. Hoffman, 1994; E. K. Hoffman, 1998.) In fact, just letting subjects know that their choices are no longer monitored by experimenters has the same effect on their behavior (Mironova, 2013).

[7] This applies to exercises in which I ask students for their predictions without explicitly stating assumptions.

 DOI: 10.7206/DEC.1733-0092.44

The weight they put on their own payment is a measure of their egoism and the weight they put on the payment of the other player is a measure of their altruism. When $w = 0$ Mary cares only about what Stan gets (she is a pure altruist), when $w = 1$ Mary is a pure egoist as defined by A1, and when $w = ½$ Mary cares equally about her and Stan's share of the division. It will be instructive to consider predictions of this model.

While predictions of this model are obtained using the same reasoning that was used before, since this case uses a different version of A1 it may be prudent to restate this reasoning again. The general solution for $w > ½$ follows from the following reasoning: Assume that the weight both players put on their own payoff is w, where $½ < w \leq 1$; the weight they put on the payoff of the other player is $1 - w$. Moreover, suppose that they both use the same weight and the weight they use is common knowledge. So, if the game goes to the second step, the utility of (0, 90) to Stan is $w90 + (1 - w)0 = 90w$ while any other division (Y, 90 – Y) where $Y > 0$ has a utility of $w(90 - Y) + (1 - w)Y = 90w + Y(1 - 2w)$ which is smaller than $90w$ since $(1 - 2w) < 0$ what is equivalent to the assumption that $w > ½$. Thus, if the game goes to the second stage Stan will offer (0, 90). Note that Mary will accept this offer for any $w < 1$ since its utility for her is $w0 + (1 - w)90 = (1 - w)90$ which will be larger than the utility of an alternative (0, 0) which is 0. When $w = 1$ Mary is indifferent between accepting and rejecting Stan's offer since her utility is 0 in both cases. Thus we need to ask if Mary can get more than $90(1 - w)$ by proposing a division that would be worth to Stan more than (0, 90). Since Stan's utility of (0, 90) is $90w$, the division of 100, denote it by (X, 100 – X), that would give him greater utility has to satisfy $w(100 - X) + (1 - w)X > w90$ which implies that $X = \lfloor 10w/(2w - 1) \rfloor$ where $\lfloor a \rfloor$ denotes a floor function which for any number a returns the largest integer not greater than a. This division gives Mary utility of $w\lfloor 10w/(2w - 1) \rfloor + (1 - w)[100 - \lfloor 10w/(2w - 1) \rfloor]$ which can be shown to be larger than $(1 - w)90$ when $w > ½$. Hence $\lfloor 10w/(2w - 1) \rfloor$, $100 - \lfloor 10w/(2w - 1 \rfloor)$ gives Mary the maximal utility she can get in this game. This is the offer she will make in the first step and Stan will accept it.

For convenience, I will refer to the class of games described and analyzed above as weighted games. It is instructive to look at predictions we get for different values of w (I chose the values of w that result in some even divisions):

There are a number of interesting observations one can make about this table. I will mention two. First note that the purely egalitarian division of (50, 50) is very unstable in that a small change in the value of w can change the prediction substantially. For instance, a change in w from 0.57 to 0.5263 will change the prediction from (40, 60) to (100, 0). This is a very unintuitive property of an otherwise intuitive model. In fact, the very existence of the (100, 0) prediction for values of w between 0.5 and 0.5263

 DOI: 10.7206/DEC.1733-0092.44

is itself an interesting and unintuitive fact. Another interesting observation concerns the prediction for $w = 0.5294$. While this value of w is by no means purely altruistic, the (90, 10) the prediction it generates is exactly the same as the prediction for the purely altruistic individuals who put all the weight on the other player's share and none on their own (Game 6.) This, again, is a rather unintuitive observation.

Table 1
Weighted games

w (weight on own payoff)	Prediction (Mary's payoff first)
1 (pure egoism)	(10, 90)
0.740	(15, 85)
0.660	(20, 80)
0.597	(30, 70)
0.570	(40, 60)
0.555	(50, 50)
0.545	(60, 40)
0.538	(70, 30)
0.533	(80, 20)
0.5294	(90, 10)
0.5263	(100, 0)
$0.5 < w < 0.5263$	(100, 0)

This case completes the list of games I propose to look at as we go back to the main consideration, namely, "which of my students' predictions should I consider to be rational?" I must admit that one reason I decided to write this paper is because my initial, admittedly careless, judgment of some predictions has been wrong. Since I believe that the same error of judgment may be more common, I thought it may be useful to lay out my own mistakes to prevent others from making the same ones.

6. Observed behaviors and types of rationality

Before I offer some conclusions from my series of exercises it will be good to take a summary snapshot of what we have done. Table 2 presents the summary of the main points I will use in my conclusion.

The logical structure of what I have done in my classroom exercises and what behavioral game theorists do in scientifically proper studies is, essentially, the same: First, we have some notion of rationality, defined by a specific set of assumptions (this is often done tacitly, hence it was important for me to make it overt and explicit.)

 DOI: 10.7206/DEC.1733-0092.44

Second, we check what follows from these assumptions deductively (derive an equilibrium of a game.) And third, we look at choices made in behavioral studies and compare them with theoretical predictions.

Table 2
Assumptions, predictions and the corresponding types of rationality

		THE NOTION OF RATIONALITY REQUIRED BY THE PREDICTION		
GAME	PREDICTION	Preferences satisfy axioms of the preference theory	Preferences satisfy axioms of the expected utility theory	Preferences satisfy the assumption of self-interest
both malevolent (Game 1)	(10, 90)	YES	NO	YES
both benevolent (Game 2)	(10, 90)	YES	NO	YES
both indifferent (Game 3)	(10, 90)	YES	YES	YES
M malevolent, S benevolent (Game 4)	(11, 89)	YES	NO	YES
M benevolent, S malevolent (Game 5)	(9, 91)	YES	NO	YES
both altruistic (Game 6)	(90, 10)	YES	YES	NO
weighted games	any division	YES	YES	NO

Let's then look at some of the outcomes I have observed in my classes and consider their meaning.

Consider, first, the case in which I describe to my students the ultimatum game and, without making assumptions like A1-A4 explicit, ask them to predict the outcome. As often happens in such cases a few of my students predict (10, 90). What can I say about this prediction? First, since I did not specify the assumptions I can only guess that a student's prediction is consistent with some sets of assumptions that he might have used in his reasoning. Table 2 tells us that it would be consistent with three different sets of assumptions that were used in Game 1, Game 2 and Game 3, of which the first two are rational in the sense of preference theory but need not be rational in the sense of expected utility theory and all satisfy the assumption of self-interest. Clearly, even from this consideration alone it is impossible for me to say in what sense the (10, 90) is a rational prediction.

The problem of judging rationality gets worse when someone predicts (11, 89) or (9, 91). Such predictions are rare but I see them in every class of about 30 students. What are the chances that a student who made such prediction had actually used the esoteric assumptions of Game 4 (malevolent Mary and benevolent Stan) and Game 5 (benevolent Mary and malevolent Stan)? But if he did not use these assumptions, what assumptions did he use? Did the assumptions he used satisfy any notion of rationality and, if so, which? Did he or didn't he make a mistake when deriving this prediction?

 DOI: 10.7206/DEC.1733-0092.44

And what if a student predicts (90, 10)? These are singular cases and they don't happen in every class but I do see them with some regularity. Should we conclude that he is rational in the sense of the expected utility theory, hence also in the sense of preference theory, but violates the assumption of self-interest (see Table 1 for a possible rational explanation of (90, 10))? Or, rather, did he make some gross inferential mistake?

Finally, are students who choose, say, (60, 40) or (50, 50) rational but not self-interested as the model of weighted games would suggest, or are they not?

To address some of the questions I have just posed I often make assumptions of the game explicit and then ask subjects to make predictions under a specific set of assumptions. I have done it with Games 1 and 2 on pretty regular basis and found out that specifying the assumptions does not affect predictions significantly – I still get a majority of predictions around the egalitarian split. What can we conclude from this observation?[8] Note that in this case I ask students to consider the case of rational players, say, in the sense of Game 1 (both malevolent.) Hence if their prediction is different from the equilibrium of this game, is it because they have failed to incorporate one the assumptions of rationality and in this sense they have failed to be rational or did they rather follow the assumptions of rationality but made a deductive mistake in their reasoning? There is no way for me to say in what way, if any, they were rational merely by looking at the prediction.

In the opening paragraph I have listed a set of statements that are often used to describe results in behavioral game theory. They were: "People are not rational." "People are not self-interested." "Real choices are (very) different from game theoretic predictions." I have often been tempted to make such judgments myself and I have often made them. It is only when I sat down one day and decided to address the problem of rationality in a more careful way did I decide I cannot conclude any of what I thought I could conclude from the casual data I have observed. Of course, in my classroom exercises I have never meant to test such conditions. But if I ever did, the task of setting up a properly controlled set of experiments would be way too complex for me to consider. Judging rationality of choice in the simple game I have described here may, in fact, be prohibitively difficult for anyone to test. But if so, I would never know if students who solve the game I present them with are rational or not? Would I?

[8] One might be tempted to conclude that such predictions simply prove that people are unable to make inferences of certain complexity. (In fact, the infamous Wason's experiment (Wason, 1968) seems to suggest that people are unable to make some very basic inferences.) But, ultimately, we don't know. One thing we know for sure is that people make notorious inferential mistakes. This has been shown in countless studies across many different disciplines and subject matters. If we take this observation as given then whenever we give people any cognitive task which is short of trivial we can bet on one thing: most will fail to solve it properly.

 DOI: 10.7206/DEC.1733-0092.44

Bibliography

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., List, J. A. (2011). Stakes matter in ultimatum games. *The American Economic Review, 101*(7), 3427-3439.

Camerer, F. C. (2011). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton NJ: Princeton University Press.

Camerer, C., Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives, 9*(2), 209-219.

Cameron, L. A. (1999). Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry, 37*(1), 47-59.

Davis, D. D., Holt, Ch. A. (1993). *Experimental Economics*. Princeton: Princeton University Press.

Güth, W., Schmittberger, R., Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization, 3*(4), 367-388.

Hoffman, E., McCabe, K. A., Smith, V. L. (1998). Behavioral foundations of reciprocity: Experimental economics and evolutionary psychology. *Economic Inquiry*, *36*(3), 335-352.

Hoffman, E., McCabe, K., Shachat, K., Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior, 7*(3), 346-380.

Kaminski, M., Lissowski, G., Swistak, P. (2013). Formal theory and value judgments. *Polish Sociological Review, 4*(184), 409-429.

Levitt, S. D., List, J. A., Sadoff, S. E. (2011). Checkmate: Exploring backward induction among chess players. *The American Economic Review, 101*(2), 975-90.

McKelvey, R. D., Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior, 1*0, 6-38.

Mironova, V., Whitt, S. (2013). *International peacekeeping, and positive peace: Experimental and survey evidence from Kosovo*. Available at SSRN 2243770, 2013.

Palacios-Huerta, I., Volij, O. (2009). *Field centipedes*. *The American Economic Review, 99*(4), 1619-35.

Slonim, R., Roth, A. E. (1998). Learning in high stakes ultimatum games: An experiment in the Slovak Republic. *Econometrica, 66*(3), 569-596.

Wason, P. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology, 20*(3), 273-281.

Weibull, J. (2004). Testing Game Theory. In: *Advances in Understanding Strategic Behaviour: Game Theory, Experiments and Bounded Rationality* (pp. 85-104). Essay in Honour of Werner Güth, by Steffen Huck. Basingstoke, UK: Palgrave MacMillan.

 DOI: 10.7206/DEC.1733-0092.44