THE TRAVELER'S

By Kaushik Basu

Lucy and Pete,
returning from a remote Pacific island,

find that the airline has damaged the identical antiques that each had purchased. An airline manager says that he is happy to compensate them but is handicapped by being clueless about the value of these strange objects. Simply asking the travelers for the price is hopeless, he figures, for they will inflate it.

Instead he devises a more complicated scheme. He asks each of them to write down the price of the antique as any dollar integer between 2 and 100 without conferring together. If both write the same number, he will take that to be the true price, and he will pay each of them that amount. But if they write different numbers, he will assume that the lower one is the actual price and that the person writing the higher number is cheating. In that case, he will pay both of them the lower number along with a bonus and a penalty—the person who wrote the lower number will get $2 more as a reward for honesty and the one who wrote the higher number will get $2 less as a punishment. For instance, if Lucy writes 46 and Pete writes 100, Lucy will get $48 and Pete will get $44.
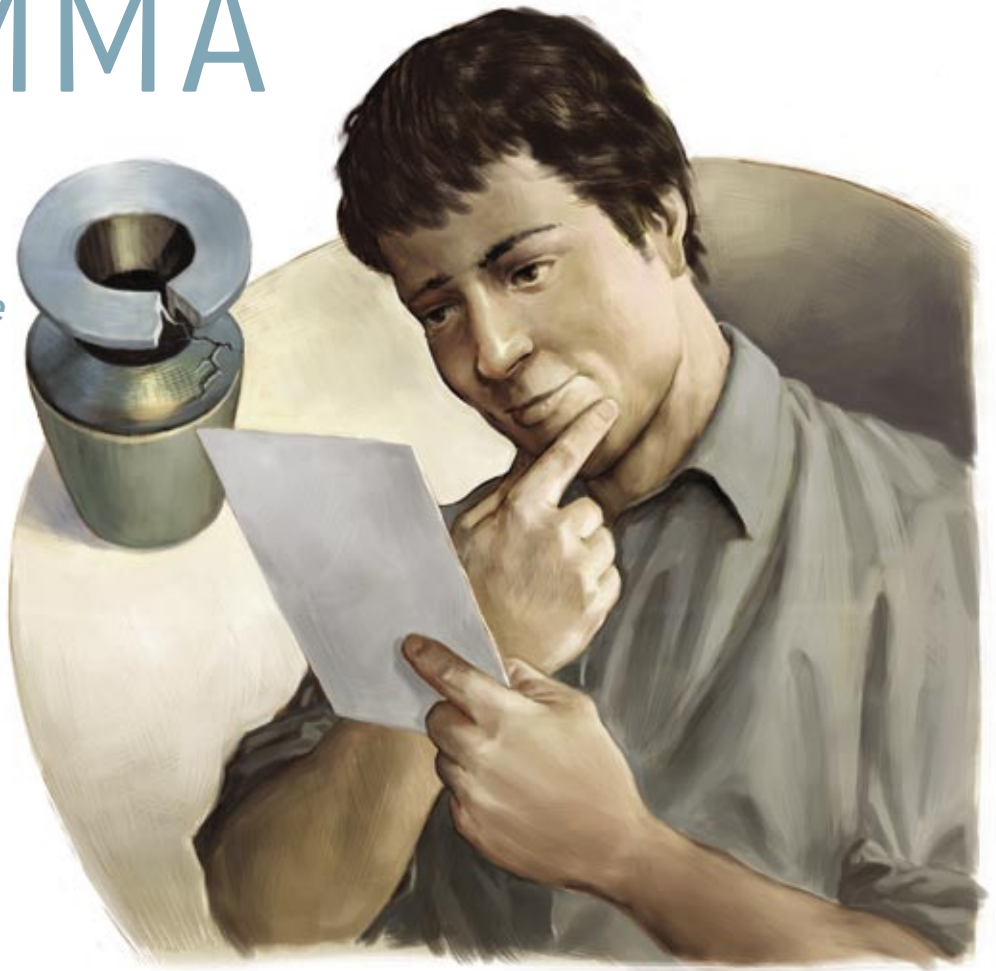
What numbers will Lucy and Pete write? What number would you write?

Scenarios of this kind, in which one or more individuals have choices to make and will be rewarded according to those choices, are known as games by the people who study them (game theorists). I crafted this game, "Traveler's Dilemma," in 1994 with several objectives in mind: to contest the narrow view of rational behavior and cognitive processes taken by economists and many political scientists, to challenge the libertarian presumptions of traditional economics and to highlight a logical paradox of rationality.

Traveler's Dilemma (TD) achieves those goals because the game's logic dictates that 2 is the best option, yet most people pick 100 or a number close to 100—both those who have not thought through the logic and those who fully understand that they are deviating markedly from the "rational" choice. Furthermore, players reap a greater reward by not adhering to reason in this way. Thus, there is something rational about

MATT COLLINS

DILEMMA

When playing this simple game, people consistently reject the rational choice. In fact, by acting illogically, they end up reaping a larger reward—an outcome that demands a new kind of formal reasoning

choosing not to be rational when playing Traveler's Dilemma.

In the years since I devised the game, TD has taken on a life of its own, with researchers extending it and reporting findings from laboratory experiments. These studies have produced insights into human decision making. Nevertheless, open questions remain about how logic and reasoning can be applied to TD.

Common Sense and Nash

TO SEE WHY 2 is the logical choice, consider a plausible line of thought that Lucy might pursue: her first idea is that she should write the largest possible number, 100, which will earn her $100 if Pete is similarly greedy. (If the antique actually cost her much less than $100, she would now be happily thinking about the foolishness of the airline manager's scheme.)

Soon, however, it strikes her that if she wrote 99 instead, she would make a little more money, because in that case she would get $101. But surely this insight will also occur to Pete, and if both wrote 99, Lucy would get $99. If Pete wrote 99, then she could do better by writing 98, in which case she would get $100. Yet the same logic would lead Pete to choose 98 as well. In that case, she could deviate to 97 and earn $99. And so

on. Continuing with this line of reasoning would take the travelers spiraling down to the smallest permissible number, namely, 2. It may seem highly implausible that Lucy would really go all the way down to 2 in this fashion. That does not matter (and is, in fact, the whole point)—this is where the logic leads us.

Game theorists commonly use this style of analysis, called backward induction. Backward induction predicts that each player will write 2 and that they will end up getting $2 each (a result that might explain why the airline manager has done so well in his corporate career). Virtually all models used by game theorists predict this outcome for TD—the two players earn $98 less than they would if they each naively chose 100 without thinking through the advantages of picking a smaller number.

Traveler's Dilemma is related to the more popular Prisoner's Dilemma, in which two suspects who have been arrested for a serious crime are interrogated separately and each has the choice of incriminating the other (in return for leniency by the authorities) or maintaining silence (which will leave the police with inadequate evidence for a case, *if* the other prisoner also stays silent). The story sounds very different from our tale of two travelers with damaged souvenirs, but the mathematics of the rewards for each option in Prisoner's

Dilemma is identical to that of a variant of TD in which each player has the choice of only 2 or 3 instead of every integer from 2 to 100.

Game theorists analyze games without all the trappings of the colorful narratives by studying each one's so-called payoff matrix—a square grid containing all the relevant information about the potential choices and payoffs for each player [*see box on opposite page*]. Lucy's choice corresponds to a row of the grid and Pete's choice to a column; the two numbers in the selected square specify their rewards.

Despite their names, Prisoner's Dilemma and the two-choice version of Traveler's Dilemma present players with no real dilemma. Each participant sees an unequivocal correct choice, to wit, 2 (or, in the terms of the prisoner story line, incriminate the other person). That choice is called the dominant choice because it is the best thing to do no matter what the other player does. By choosing 2 instead of 3, Lucy will receive $4 instead of $3 if Pete chooses 3, and she will receive $2 instead of nothing if Pete chooses 2.
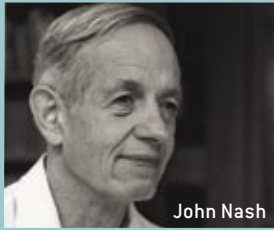
$101. Because Lucy is better off by this change, the outcome (100, 100) is not a Nash equilibrium.

TD has only one Nash equilibrium—the outcome (2, 2), whereby Lucy and Pete both choose 2. The pervasive use of the Nash equilibrium is the main reason why so many formal analyses predict this outcome for TD.

Game theorists do have other equilibrium concepts—strict equilibrium, the rationalizable solution, perfect equilibrium, the strong equilibrium and more. Each of these concepts leads to the prediction (2, 2) for TD. And therein lies the trouble. Most of us, on introspection, feel that we would play a much larger number and would, on average, make much more than $2. Our intuition seems to contradict all of game theory.

Implications for Economics

THE GAME AND OUR INTUITIVE prediction of its outcome also contradict economists' ideas. Early economics was firmly tethered to the libertarian presumption that individuals should be left to their own devices because their selfish choices will re-

> # Game theory predicts that the Nash equilibrium will occur when Traveler's Dilemma is played rationally.

John Nash

In contrast, the full version of TD has no dominant choice. If Pete chooses 2 or 3, Lucy does best by choosing 2. But if Pete chooses any number from 4 to 100, Lucy would be better off choosing a number larger than 2.

When studying a payoff matrix, game theorists rely most often on the Nash equilibrium, named after John F. Nash, Jr., of Princeton University. (Russell Crowe portrayed Nash in the movie *A Beautiful Mind*.) A Nash equilibrium is an outcome from which no player can do better by deviating unilaterally. Consider the outcome (100, 100) in TD (the first number is Lucy's choice, and the second is Pete's). If Lucy alters her selection to 99, the outcome will be (99, 100), and she will earn

sult in the economy running efficiently. The rise of game-theoretic methods has already done much to cut economics free from this assumption. Yet those methods have long been based on the axiom that people will make selfish rational choices that game theory can predict. TD undermines both the libertarian idea that unrestrained selfishness is good for the economy and the game-theoretic tenet that people will be selfish and rational.

In TD, the "efficient" outcome is for both travelers to choose 100 because that results in the maximum total earnings by the two players. Libertarian selfishness would cause people to move away from 100 to lower numbers with less efficiency in the hope of gaining more individually.

And if people do not play the Nash equilibrium strategy (2), economists' assumptions about rational behavior should be revised. Of course, TD is not the only game to challenge the belief that people always make selfish rational choices [see "The Economics of Fair Play," by Karl Sigmund, Ernst Fehr and Martin A. Nowak; SCIENTIFIC AMERICAN, January 2002]. But it makes the more puzzling point that even if players have no concern other than their own profit, it is not rational for them to play the way formal analysis predicts.

TD has other implications for our understanding of real-world situations. The game sheds light on how the arms race acts as a gradual process, taking us in small steps to ever worsening outcomes. Theorists have also tried to extend TD to understand how two competing firms may undercut each other's price to their own detriment (though in this case to

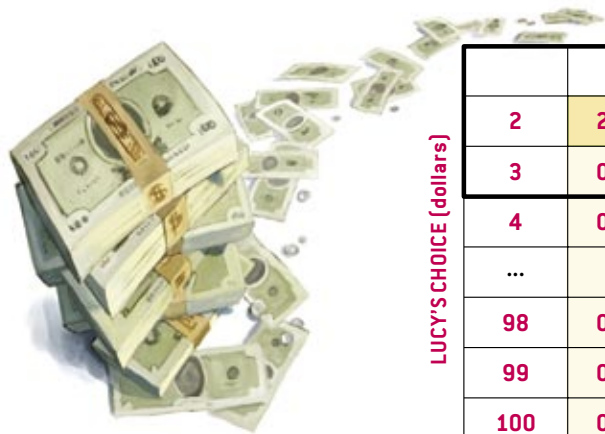# Overview/*Sensible Irrationality*

- In the game Traveler's Dilemma, two people separately choose a whole number from 2 to 100, and the player whose number is smaller is rewarded with a larger sum. Game theory insists that rationality should lead players to select 2, but most people pick an integer closer to 100.
- A new kind of reasoning is needed to gain a rigorous understanding of this rational choice not to be rational.
- The results of Traveler's Dilemma contradict economists' assumption that standard game theory can predict how supposedly selfish rational people will behave. They also show how selfishness is not always good economics.

PAYOFF MATRIX OF THE TRAVELER'S DILEMMA

This payoff matrix summarizes everything game theorists need to know about Traveler's Dilemma. Lucy's possible choices are shown in the leftmost column; Pete's run across the top row. The first number in the square at the intersection of the chosen row and column is Lucy's payoff, and the second number is Pete's payoff. For example, if Lucy chooses 98 and Pete 99, then Lucy receives $100 and Pete receives $96.

The outcome in which both players choose 2 and each earns $2 (*gold*) is called the Nash equilibrium: Lucy does worse (earning $0) if she chooses any other number and Pete still chooses 2. Similarly, Pete does worse if he alone chooses something other than 2.

When restricted to choices of only 2 and 3 (*black outline*), the game becomes equivalent to Prisoner's Dilemma.

PETE'S CHOICE (dollars)

LUCY'S CHOICE (dollars)	2	3	4	...	98	99	100
2	2 2	4 0	4 0	...	4 0	4 0	4 0
3	0 4	3 3	5 1	...	5 1	5 1	5 1
4	0 4	1 5	4 4	...	6 2	6 2	6 2
...	...	...	...	...	...	...	...
98	0 4	1 5	2 6	...	98 98	100 96	100 96
99	0 4	1 5	2 6	...	96 100	99 99	101 97
100	0 4	1 5	2 6	...	96 100	97 101	100 100

the advantage of the consumers who buy goods from them).

All these considerations lead to two questions: How do people actually play this game? And if most people choose a number much larger than 2, can we explain why game theory fails to predict that? On the former question, we now know a lot; on the latter, little.

How People Actually Behave

OVER THE PAST DECADE researchers have conducted many experiments with TD, yielding several insights. A celebrated lab experiment using real money with economics students as the players was carried out at the University of Virginia by C. Monica Capra, Jacob K. Goeree, Rosario Gomez and Charles A. Holt. The students were paid $6 for participating and kept whatever additional money they earned in the game. To keep the budget manageable, the choices were valued in cents instead of dollars. The range of choices was made 80 to 200, and the value of the penalty and reward was varied for different runs of the game, going as low as 5 cents and as high as 80 cents. The experimenters wanted to see if varying the magnitude of the penalty and reward would make a difference in how the game was played. Altering the size of the reward and penalty does not change any of the formal analysis: backward induction always leads to the outcome (80, 80), which is the Nash equilibrium in every case.

The experiment confirmed the intuitive expectation that the average player would not play the Nash equilibrium strategy of 80. With a reward of 5 cents, the players' average choice was 180, falling to 120 when the reward rose to 80 cents.

Capra and her colleagues also studied how the players' behavior might alter as a result of playing TD repeatedly.

Would they learn to play the Nash equilibrium, even if that was not their first instinct? Sure enough, when the reward was large the play converged, over time, down toward the Nash outcome of 80. Intriguingly, however, for small rewards the play increased toward the opposite extreme, 200.

The fact that people mostly do not play the Nash equilibrium received further confirmation from a Web-based experiment with no actual payments that was carried out by Ariel Rubinstein of Tel Aviv University and New York University from 2002 to 2004. The game asked players, who were going to attend one of Rubinstein's lectures on game theory and Nash, to choose an integer between 180 and 300, which they were to think of as dollar amounts. The reward/penalty was set at $5.

Around 2,500 people from seven countries responded, giving a cross-sectional view and sample size infeasible in a laboratory. Fewer than one in seven players chose the scenario's Nash equilibrium, 180. Most (55 percent) chose the maximum number, 300 [*see box on next page*]. Surprisingly, the data were very similar for different subgroups, such as people from different countries.

The thought processes that produce this pattern of choices remain mysterious, however. In particular, the most popular response (300) is the only strategy in the game that is

THE AUTHOR

KAUSHIK BASU is professor of economics, Carl Marks Professor of International Studies and director of the Center for Analytic Economics at Cornell University. He has written extensively in academic journals on development economics, welfare economics, game theory and industrial organization. He also writes for the popular media, including a monthly column in BBC News online. He is a fellow of the Econometric Society.

MATT COLLINS

"dominated"—which means there is another strategy (299) that never does worse and sometimes does better.

Rubinstein divided the possible choices into four sets of numbers and hypothesized that a different cognitive process lies behind each one: 300 is a spontaneous emotional response. Picking a number between 295 and 299 involves strategic reasoning (some amount of backward induction, for instance). Anything from 181 to 294 is pretty much a random choice. And finally, standard game theory accounts for the choice of 180, but players might have worked that out for themselves or may have had prior knowledge about the game.

A test of Rubinstein's conjecture for the first three groups would be to see how long each player took to make a decision. Indeed, those who chose 295 to 299 took the longest time on average (96 seconds), whereas both 181 to 294 and 300 took about 70 seconds—a pattern that is consistent with his hypothesis that people who chose 295 to 299 thought more than those who made other choices.

Game theorists have made a number of attempts to explain why a lot of players do not choose the Nash equilibrium in TD experiments. Some analysts have argued that many people are unable to do the necessary deductive reasoning and therefore make irrational choices unwittingly. This explanation must be true in some cases, but it does not account for all the results, such as those obtained in 2002 by Tilman Becker, Michael Carter and Jörg Naeve, all then at the University of Hohenheim in Germany. In their experiment, 51 members of the Game Theory Society, virtually all of whom are professional game theorists, played the original 2-to-100 version of TD. They played against each of their 50 opponents by selecting a strategy and sending it to the researchers. The strategy could be a single number to use in every game or a selection of numbers and how often to use each of them. The game had a real-money reward system: the experimenters would select one player at random to win $20 multiplied by that player's average payoff in the game. As it turned out, the winner, who had an average payoff of $85, earned $1,700.

Of the 51 players, 45 chose a single number to use in every game (the other six specified more than one number). Among those 45, only three chose the Nash equilibrium (2), 10 chose the dominated strategy (100) and 23 chose numbers ranging from 95 to 99. Presumably game theorists know how to reason deductively, but even they by and large did not follow the rational choice dictated by formal theory.

Superficially, their choices might seem simple to explain: most of the participants accurately judged that their peers would choose numbers mainly in the high 90s, and so choosing a similarly high number would earn the maximum average return. But why did everyone expect everyone else to choose a high number?

Perhaps altruism is hardwired into our psyches alongside selfishness, and our behavior results from a tussle between the two. We know that the airline manager will pay out the largest amount of money if we both choose 100. Many of us do not feel like "letting down" our fellow traveler to try to earn only an additional dollar, and so we choose 100 even though we fully understand that, rationally, 99 is a better choice for us as individuals.
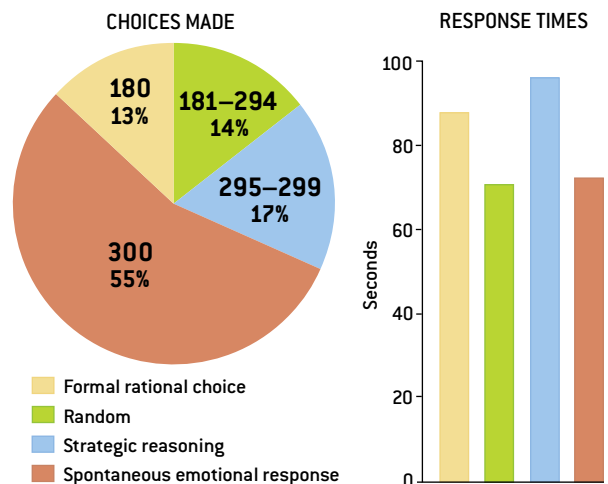
To go further and explain more of the behaviors seen in experiments such as these, some economists have made strong and not too realistic assumptions and then churned out the observed behavior from complicated models. I do not believe that we learn much from this approach. As these models and assumptions become more convoluted to fit the data, they provide less and less insight.

Unsolved Problem

THE CHALLENGE THAT REMAINS, however, is not explaining the real behavior of typical people presented with TD. Thanks in part to the experiments, it seems likely that altruism, socialization and faulty reasoning guide most individuals' choices. Yet I do not expect that many would select 2 if those three factors were all eliminated from the picture. How can we explain it if indeed most people continue to choose large numbers, perhaps in the 90s, even when they have no dearth of deductive ability, and they suppress their normal altruism and social behavior to play ruthlessly to try to make as much money as possible? Unlike the bulk of modern game theory, which may involve a lot of mathematics but is straightforward once one knows the techniques, this question is a hard one that requires creative thinking.

WHAT WERE THEY THINKING?

Researchers hypothesized that different thought processes lay behind different types of choices that people made playing a version of Traveler's Dilemma with the options ranging from 180 to 300 (*pie chart*): a spontaneous emotional response (*choosing 300*), a strategically reasoned choice (*295–299*) or a random one (*181–294*). Players making the formal rational choice (*180*) might have deduced it or known about it in advance. As expected, people making "spontaneous" or "random" selections took the least time to choose (*bar graph*).

CHOICES MADE

- 180 13%
- 181–294 14%
- 295–299 17%
- 300 55%

RESPONSE TIMES

Legend:
- Formal rational choice
- Random
- Strategic reasoning
- Spontaneous emotional response

JEN CHRISTIANSEN

Game Theory vs. Ordinary Decision Theory: I Know that You Know that I Know …

I heard this tale in India. A hat seller, on waking from a nap under a tree, found that a group of monkeys had taken all his hats to the top of the tree. In exasperation he took off his own hat and flung it to the ground. The monkeys, known for their imitative urge, hurled down the hats, which the hat seller promptly collected.

Half a century later his grandson, also a hat seller, set down his wares under the same tree for a nap. On waking, he was dismayed to discover that monkeys had taken all his hats to the treetop. Then he remembered his grandfather's story, so he threw his own hat to the ground. But,

mysteriously, none of the monkeys threw any hats, and only one monkey came down. It took the hat on the ground firmly in hand, walked up to the hat seller, gave him a slap and said, "You think only you have a grandfather?"

This story illustrates an important distinction between ordinary decision theory and game theory. In the latter, what is rational for one player may depend on what is rational for the other player. For Lucy to get her decision right, she must put herself in Pete's shoes and think about what he must be thinking. But he will be thinking about what she is thinking, leading to an infinite regression. Game theorists describe this situation by saying that "rationality is common knowledge among the players." In other words, Lucy and Pete are rational, they each know that the other is rational, they each know that the other knows, and so on.

The assumption that rationality is common knowledge is so pervasive in game theory that it is rarely stated explicitly. Yet it can run us into problems. In some games that are played over time, such as repeated rounds of Prisoner's Dilemma, players can make moves that are incompatible with this assumption.

I believe that the assumption that rationality is common knowledge is the source of the conflict between logic and

intuition and that, in the case of Traveler's Dilemma, the intuition is right and awaiting validation by a better logic. The problem is akin to what happened in early set theory. At that time, mathematicians took for granted the existence of a universal set—a set that contained everything. The universal set seemed extremely natural and obvious, yet ultimately several paradoxes of set theory were traced to the assumption that it existed, which mathematicians now know is flawed. In my opinion, the common knowledge of rationality assumed by game theorists faces a similar demise.   —K.B.

Suppose you and I are two of these smart, ruthless players. What might go through our minds? I expect you to play a large number—say, one in the range from 90 to 99. Then I should not play 99, because whichever of those numbers you play, my choosing 98 would be as good or better for me. But if you are working from the same knowledge of ruthless human behavior as I am and following the same logic, you will also scratch 99 as a choice—and by the kind of reasoning that would have made Lucy and Pete choose 2, we quickly eliminate every number from 90 to 99. So it is not possible to make the set of "large numbers that ruthless people might logically choose" a well-defined one, and we have entered the philosophically hard terrain of trying to apply reason to inherently ill-defined premises.

If I were to play this game, I would say to myself: "Forget game-theoretic logic. I will play a large number (perhaps 95), and I know my opponent will play something similar and both of us will ignore the rational argument that the next smaller number would be better than whatever number we choose." What is interesting is that this rejection of formal rationality and logic has a kind of meta-rationality attached

to it. If both players follow this meta-rational course, both will do well. The idea of behavior generated by rationally rejecting rational behavior is a hard one to formalize. But in it lies the step that will have to be taken in the future to solve the paradoxes of rationality that plague game theory and are codified in Traveler's Dilemma.   SA

MORE TO EXPLORE

On the Nonexistence of a Rationality Definition for Extensive Games. Kaushik Basu in *International Journal of Game Theory,* Vol. 19, pages 33–44; 1990.

The Traveler's Dilemma: Paradoxes of Rationality in Game Theory. Kaushik Basu in *American Economic Review,* Vol. 84, No. 2, pages 391–395; May 1994.

Anomalous Behavior in a Traveler's Dilemma? C. Monica Capra et al. in *American Economic Review,* Vol. 89, No. 3, pages 678–690; June 1999.

The Logic of Backwards Inductions. G. Priest in *Economics and Philosophy,* Vol. 16, No. 2, pages 267–285; 2000.

Experts Playing the Traveler's Dilemma. Tilman Becker et al. Working Paper 252, Institute for Economics, Hohenheim University, 2005.

Instinctive and Cognitive Reasoning. Ariel Rubinstein. Available at arielrubinstein.tau.ac.il/papers/Response.pdf

MATT COLLINS