Response Time and Decision Making:
A "Free" Experimental Study

Ariel Rubinstein*

Tel Aviv University
and
New York University

First draft: Dec. 2012

Abstract

Response time is used to interpret choice in decision problems. It is first establishes that there is a correlation between short response time and choices that are clearly a mistake. It is then determines whether a correlation also exists between response time and behavior that is inconsistent with some standard theories of decision making. The lack of such a correlation is interpreted to imply that such behavior does not reflect a mistake. It is also shown that a typology of slow and fast responders may, in some cases, be more useful than standard typologies.

Key words: Response Time, Reaction Time, Decision problems, Allais Paradox, mistakes, Neuro-economics.

1

This paper was written with the recognition that it is unlikely to be accepted for publication (although I would be pleased if it is…). There is something liberating about writing a paper without trying to please referees and without having to take into consideration the various protocols and conventions imposed on researchers in experimental economics (see Rubinstein (2001)). It gives one a feeling of real academic freedom!

Over the last 10 years, I have been gathering data on response time in web-based experiments on my didactic website gametheory.tau.ac.il. The basic motivation can be stated quite simply: response time can provide insights into the process of deliberation prior to making a decision. This is a very old idea in the psychological literature and Sternberg (1969) attributes it to a paper published by Franciscus Cornelis Donders in 1868 (an English translation appears in Donders (1969)). However, psychologists have typically analyzed response time in situations where subjects respond to visual or vocal signals and response time is measured in fractions of a second.

I am interested in a different type of problem that involves choice and strategic decisions. My focus is on response time as an indicator of the type of analyses carried out by a decision maker. Essentially, I adopt the point of view advocated for many years by Daniel Kahneman (see Kahneman (2011)). In Rubinstein (2006), I presented some experimental game theoretic results which were consistent with the distinction between instinctive strategies (which are the outcome of activating system I in Kahneman's terminology) and cognitive strategies (which are the outcome of activating system II). Thus, for example, I argued that offering 50% of the pie in an ultimatum game is the instinctive action while offering less than that is the outcome of a longer cognitive process. In the meantime, a few other papers have presented experimental results using response time as an indicator.[1]

The nature of response time as measured in the current study is different from that in the psychological literature mentioned above. Here, response time is a particularly noisy variable

---

[1] Arad and Rubinstein (2012) use response time to support the interpretation of the participants' choices as reflecting multi-dimensional iterative reasoning; Agranov, Caplin, and Tergiman (2010) use response time to interpret choice in guessing games; Branas-Garza, Meloso, and Miller (2012) study response time in the ultimatum game; Chabris, Laibson, Morris, Schuldt and Taubinsky (2008) study the correlation between response time and time preferences; Lotito, Migheli and Ortona (2011) argue, using a public goods game, that response time data supports the intuition that cooperation is instinctive; Piovesan and Wengström (2009) argue that faster subjects more often choose the option with the highest payoff for themselves; Schotter and Trevino (2012) use response time to predict behavior in global games; Studer and Winkelmann (2012) argue that slower responses and greater cognitive effort reduce reported happiness.

with a large range. In my experience, one needs hundreds of subjects responding to a particular question in order to obtain clear cut results. One of the advantages of using a website as a platform is the ability to gather large amounts of data at a low cost. Since gametheory.tau.ac.il went online (in 2002), responses have been collected from more than 45,000 students in 717 classes in 46 countries.[2]

How does the site work? Teachers who register on the site can assign problems to their students from a bank of decision and game situations. Students are promised that their responses will be kept anonymous. Recording response time started soon after the site began operating. The result is a huge data set (for some of the problems the site recorded more than 10,000 responses). The aim of this paper is to share some of the results and the insights obtained.

This type of experimental research has some obvious merits: it is cheap and facilitates the participation of thousands of subjects from a population that is far more diverse than is usually the case in experiments. The behavioral results are, in my experience, not qualitatively different from those obtained by more conventional methods.

However, there has been some heated criticism of this method of experimentation. I will let the critics speak for themselves -- following is a quote from a referee report attacking Arad and Rubinstein (2012), a paper in which the data was obtained using the site and response time was used to interpret behavior:

> "….The paper discusses a survey or a questionnaire, not an economics experiment. Decisions have no monetary consequences, and so participants have no monetary incentives to choose the strategy they think it is best. Thus, it does not seem right to describe the study as an economics experiment. The second flaw is that the study does not measure participants' engagement time with the game. There is no control over what participants are doing. Participants can take a break to take a mobile phone call, to text someone, to browse the web, to eat pizza, to have a coffee, etc. etc. Yet, response times are presented as indicative of the time taken to achieve a decision. The paper completely ignores this issue. Since the time recorded goes up when people are doing other things other than thinking about the survey game they are playing, the paper does not measure participants' response time."

The referee has eloquently raised two issues:

---

[2] Argentina, Austria, Australia, Belgium, Brazil, Brunei Darussalam, Bulgaria, Canada, Chile, China, Columbia, Denmark, Ecuador, Estonia, Finland, France, Germany, Greece, Guatemala, Hong Kong, Hungary, India, Ireland, Israel, Italy, Korea, Kyrgyzstan, Malaysia, Mexico, Moldova, Netherlands, New Zealand, Norway, Peru, Poland, Portugal, the Russian Federation, Singapore, the Slovak Republic, Spain, Switzerland, Taiwan, Turkey, the UK, the US and Vietnam.

(a) The lack of monetary incentives. I have never understood the source of the myth that paying a few dollars (with some probability) will make the subjects (who come to the lab on their own volition and are paid a certain amount no matter how they perform in the experiment) as focused on the task as they would be in real life. The opposite would seem to be the case. Human beings have a good imagination and framing a question using "imagine that…" achieves a degree of focus equal at least to that created by a small monetary incentive. Exceptions might include very boring tasks in which incentives are necessary to make sure subjects are not just answering arbitrarily. In any case, I cannot see how the incentive provided by the small amount of money involved can be compared to the advantage gained from easy and quick access to a large number of subjects from a variety of countries.

(b) The referee attacks the use of a non-laboratory setting. He claims that using web-based experiments does not provide control over what participants are doing. Well…. do researchers know whether a subject in a Lab is thinking about the experiment or perhaps about his troubled love life? Are decisions more natural in a "sterile environment" or when a subject is sitting at home eating pizza? (An attendee of a lecture of mine on the subject referred me to a paragraph from "War and Peace" which describes Kutuzov deciding about the battle at Borodino which turned the tide of the war against Napoleon : "Kutuzov was chewing a piece of roast chicken with difficulty and glanced at Wolzogen with eyes that brightened under their puckering lids……").

Note also that response time will be used in this paper only as a "relative measure". In other words, I will not relate to the absolute value of the response time but will only compare the response time distributions of subjects who chose alternative A with those who chose alternative B. Therefore, unless eating pizza affects the choice itself it is not a major concern…

I will not be presenting any results of statistical testing. Given the number of observations and the clarity of the findings, statistical testing appears to be superfluous. I will draw conclusions only when there is a major difference between two populations and the number of subjects is in the hundreds. Although statistics is a fascinating subject, I feel that its use in experimental economics is often vacuous since it focuses on a particular type of uncertainty and misses the larger uncertainties related to the method of data collection, the reliability of the researcher, etc. I prefer to deal with results that lead to a crystal clear conclusion and if a statistical test puts that conclusion into doubt, then the test itself rather the conclusion should be re-considered…

In any case, I do not believe that the standard methods used in experimental economics guarantee conclusions that are more reliable or "scientific". A researcher in the social sciences must apply common sense, interpret his results (somewhat subjectively) and decide whether the results are spurious or indeed provide significant insights. I have tried to do this faithfully but I am aware that my own judgment may also be subject to biases.

4

The current paper deals only with decision problems. In a companion paper (yet to be written) I will report on results from dozens of strategic situations (i.e. games).

The main argument of the paper is structured as follows:

A. The first task is to demonstrate that response time, as measured here, confirms the intuition that there is a strong negative correlation between response time and mistakes, i.e. response time is shorter in the case of mistakes.

B. Several well-known problems that are often used to demonstrate behavior that conflicts with standard theories of decision making (such as the Allais paradox, the Ellsberg paradox and framing affects) are then examined. No correlation is found between short response time and behavior that is "inconsistent with the theory". It is concluded that "inconsistent behavior" is not similar to "making mistakes". Thus, inconsistent behavior is associated with the same level of cognitive investment as behavior that is consistent with the standard notions of rationality and thus one cannot dismiss inconsistent behavior as merely "mistaken".

C. The possibility will be explored of identifying a new typology of decision making which distinguishes between agents according to their speed of decision making rather than their standard preferences. It will be suggested that in some problems (such as the Allais paradox) a fast/slow response typology might be more useful than a more/less risk averse typology.

1. Preliminaries

Subjects: Almost all subjects were students in Game Theory courses in various countries around the world.[3] More than half of them were from the US, Switzerland, the UK, Columbia, Argentina and the Slovak Republic.

Problems: Each problem contains a short description of a situation, decision problem or game and a hypothetical question about the subject's anticipated behavior. The analysis includes 116 problems, each of which were answered by at least 600 subjects.

Response Time (RT): Response time is measured as the length of time from the moment a problem is sent to a subject until his response is received by the server (in Tel Aviv).

---

[3] A few teachers assigned to their students more than 25 problems in one set. The data of those courses was excluded. A restriction was imposed several years ago such that teachers could not assign more than 15 games in one set.

Median Response Time (MRT): One of the main tools used in the analysis is the median of subjects' response times, which is calculated for a particular choice (or class of choices) in a decision problem.

The cumulative RT distribution: For every alternative x, $F_x(t)$ is the proportion of subjects who responded within t seconds from among those who chose x.

Comparing the distributions F_x and F_y is the basic tool used to compare two responses x and y. The graphs of the cumulative RT distributions display two remarkable regularities:

(a) They all have a similar shape (when the number of participants is large), which consistently takes the form of a smooth concave function. The graphs of the distributions remind us the shape of the inversegaussian and lognormal distributions but a statistician tells me that no familiar distribution could be clearly identified with those graphs. When the number of subjects increases, the graphs become so smooth that you want to take a derivative…☺

(b) The cumulative RT distributions for the various responses to a question can be ordered by the "first order stochastic domination" relation (that is, $F_x(t) \geq F_y(t)$ for all t). This is probably the clearest evidence one can expect to find in such data that it takes longer for subjects to choose y than to choose x. The larger is the gap between the two distributions, the stronger this claim becomes.

Rankings: A subject's "**Local rank**" is the fraction of subjects who answered a particular question faster than he did. A "local rank" is calculated for each participant for each of the questions he answered. A "**Global Rank**" is attached to each participant who answered at least 9 problems. The median number of answers for a participant who received a global rank is 15. The "global rank**"** of a subject is the median of his local rankings. About 50% of the subjects received this global rank. The following graph presents the distribution of global rankings:

The distribution of global rankings
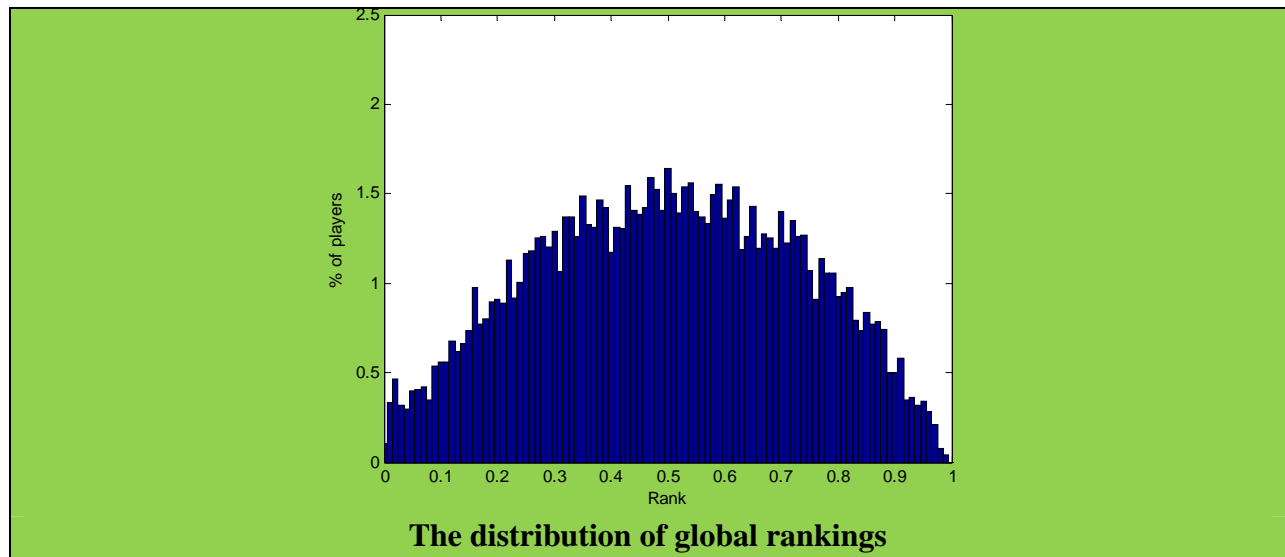
This distribution of global rankings is much more dispersed than obtained if the local rankings were distributed uniformly (see http://www.sjsu.edu/faculty/watkins/samplemedian.htm). This indicates the existence of a significant correlation between the local rankings of the same individual in different problems.

Correlation between the local and global rankings: The correlation between local and global ranks was calculated for each problem. With one exception, it always lies within the interval (0.5, 0.7) and is usually in the vicinity of 0.6.

n and m: For each problem, n and m are defined as follows:
  n - the total number of subjects who responded to a question.
  m - the total number of subjects who responded to a question and have a global rank.

Tables: The following are the standard tables used in the paper:

For a single problem:
a. **Basic table**: presents the basic statistics of the distribution of answers and the MRT, as well as a graph of the cumulative distributions for the main responses.
b. **Slow/Fast quartiles**: The subjects responding to a particular problem are divided into quartiles according to their local and global rankings. For each quartile, the distribution of answers (and their MRT) is reported.

For a pair of problems:
In some cases, subjects responded to two related questions, in which case the following tables are presented:
c. **Basic table**: presents the joint distributions of answers.

7

d. Slow/Fast quartiles: Subjects who responded faster than the median in both problems are classified as "fast" and those who responded slower are classified as "slow". About 75% of the subjects are included in these two equally-sized groups. The two joint distributions of responses will be presented.

2. Cases in which the RT is an indicator of a mistake

This section presents four problems in which the response time results are consistent with the hypothesis that response is significantly faster for choices that clearly involve a mistake.

A. Count the F's

The "Count the F's" problem has been floating around the Internet for a number of years (I don't know its origin).

Subjects are asked to count the number of appearances of the letter F in the following 80 letters text:

FINISHED FILES ARE THE RE-
SULT OF YEARS OF SCIENTIF-
IC STUDY COMBINED WITH
THE EXPERIENCE OF YEARS.

Defining a mistake is straightforward in this case: either you count the number of F's correctly or you don't. Many readers will be surprised to learn that the right answer is 6 and not 3. In fact, only 37% of the subjects (n=5324 and m=4453) answered the question correctly. The reason for this is that many people (including me) tend to skip the word "of" when reading a text and in the above text it appears three times. Only 3% of the subjects gave an answer outside the range 3-6. This can be taken as evidence that subjects took the task more seriously than some critics would claim. The MRT results and the standard graphs present a clear picture in this case:

n=5324	Percent	MRT
3	36%	48s
4	11%	54s
5	13%	54s
6	37%	59s

Count the F's: Basic statistics
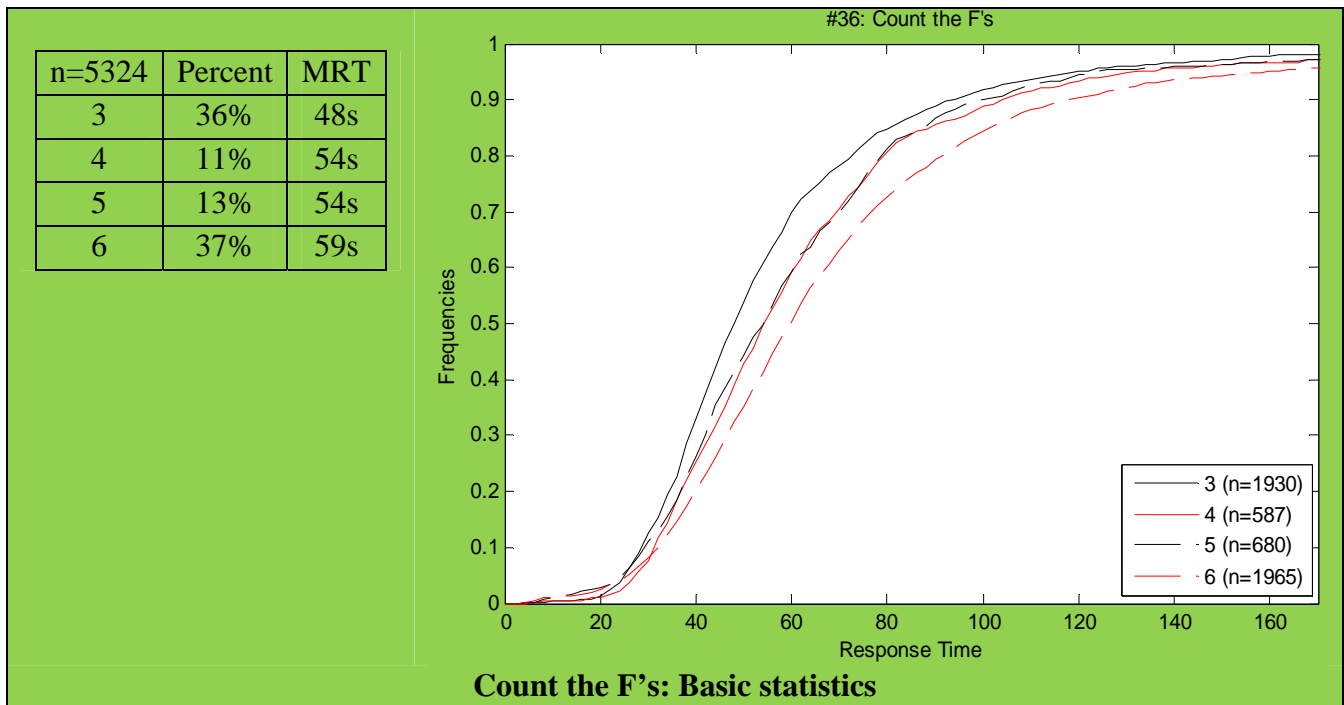
Subjects who made a large mistake (i.e. they answered 3) spent about 6s less on the problem than those who answered "4" or "5" and 11s less than those who answered "6".

Note the similarity in response times between the answers "4" and "5". In some sense, "4" is a "bigger" mistake than "5". However, note that the word "of" appears twice in the second line and once in the last line. It seems reasonable to assume that a person who notices one of the "of"'s in the second line will also notice the other. Thus, the "size" of the mistake is in fact similar for the subjects who chose "4" or "5" and the response times support this view.

	Local ranking quartiles				Global ranking quartiles			
Response	Fastest	Fast	Slow	Slowest	Fastest	Fast	Slow	Slowest
3	46%	42%	31%	26%	43%	36%	34%	29%
4	11%	11%	12%	11%	12%	12%	10%	10%
5	13%	13%	14%	12%	13%	12%	14%	13%
6	29%	32%	41%	46%	29%	37%	39%	45%
n	1331	1331	1331	1331	1134	1166	1145	1008
MRT	31s	46s	62s	98s	39s	50s	59s	80

Count the F's: Results according to local and global rankings

The above table presents the distribution of the answers 3,4,5 and 6 in each of the four quartiles. (The few answers below 3 and above 6 are not reported.) According to the local rankings, the proportion of "large" mistakes among the slowest quartile of subjects is 26% whereas it is 46% among the fastest. Furthermore, 46% of the slowest participants gave the correct answer as compared to only 29% of the fastest. Similar though less pronounced differences were obtained for global rankings.

B. Most Likely Sequence

Kahneman and Tversky (1983) report on the following experiment:

> Subjects were asked to consider a six-sided die with four green faces and two red ones. The subjects were told that the die will be rolled 20 times and the sequence of G and R recorded. Each subject was then asked to select one of the following three sequences and told that he would receive $25 if it appears during the rolls of the die. The three sequences were:
> RGRRR
> GRGRRR
> GRRRRR

The right answer is of course RGRRR. Only 5% of the subjects (n=2316 and m=2159) chose the blatantly incorrect answer GRRRRR, although 58% chose the other mistake GRGRRR (note that whenever this sequence appears, RGRRR appears as well). The MRT for the right answer is much higher (by 16s) than for the mistake, which is the far more common answer.

n=2316	Percent	MRT
RGRRR	37%	78s
GRGRRR	58%	62s
GRRRRR	5%	49s

Most Likely Sequence: Basic statistics
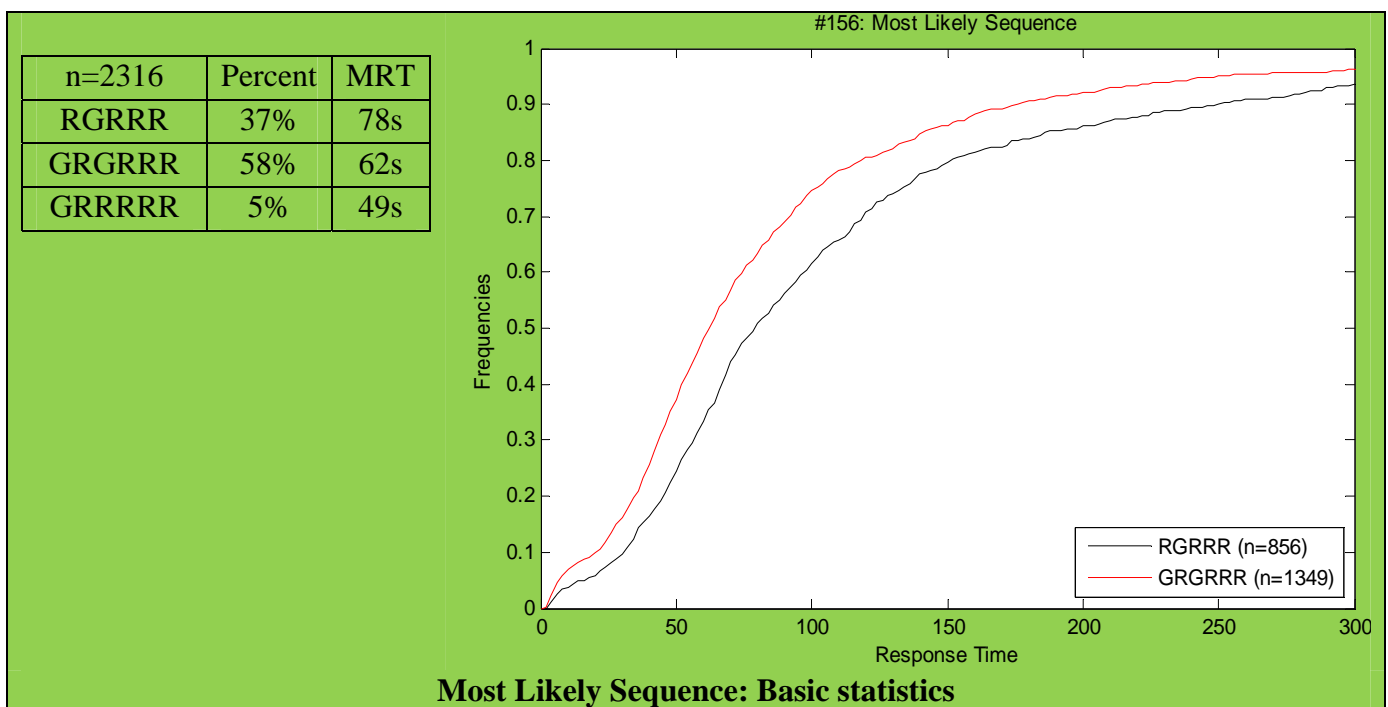
According to the local rankings, 47% of the slowest quartile of subjects gave the correct answer as compared to only 26% of the fastest quartile. Similar differences were obtained for the global rankings. The correlation between the local and global rankings is 0.64.

	Local ranking quartiles				Global ranking quartiles			
Response	Fastest	Fast	Slow	Slowest	Fastest	Fast	Slow	Slowest
RGRRR	26%	34%	40%	47%	28%	36%	42%	43%
GRGRRR	66%	62%	57%	49%	66%	59%	54%	53%
GRRRRR	8%	5%	3%	4%	6%	5%	4%	4%
n	579	579	579	579	544	560	550	505
MRT	26s	54s	85s	173s	34s	59s	80s	136s

Most Likely Sequence: Results according to local and global rankings

C. Two Roulettes

Tversky and Kahneman (1986) demonstrated the existence of framing effects using the following experiment:

Subjects are asked to (virtually) choose between two roulette games:

Roulette A	Color	White	Red	Green	Yellow
	Chances %	90	6	1	3
	Prize $	0	45	30	-15

Roulette B	Color	White	Red	Green	Yellow
	Chances %	90	7	1	2
	Prize $	0	45	-10	-15

Subjects were divided almost equally between choosing A and B (n=2785, m=2319). The choice of A is probably an outcome of using a "cancelling out" procedure in which similar parameters (Red and Yellow) are ignored and subjects focus on the large differences in the Green parameter ($30 vs -$10). However, choosing according to Green is a mistake since roulettes A and B are "identical" to the following roulettes C and D respectively (n=2245, m=1854) and D clearly "dominates" C:

Roulette C	Color	White	Red	Green	Blue	Yellow
	Chances %	90	6	1	1	2
	Prize $	0	45	30	-15	-15

Roulette D	Color	White	Red	Green	Blue	Yellow
	Chances %	90	6	1	1	2
	Prize $	0	45	45	-10	-15

The RT results dramatically confirm that making a wrong choice is correlated with shorter RT:

n=2785	Percent	MRT	(std)
A	50%	54s	(1.2s)
B	50%	93s	(2.8s)

#59: Roulettes (part 1)

n=2245	Percent	MRT	(std)
C	7%	16s	(2.5s)
D	93%	32s	(0.7s)

#60: Roulettes (part 2)

The Two Roulettes: Basic statistics
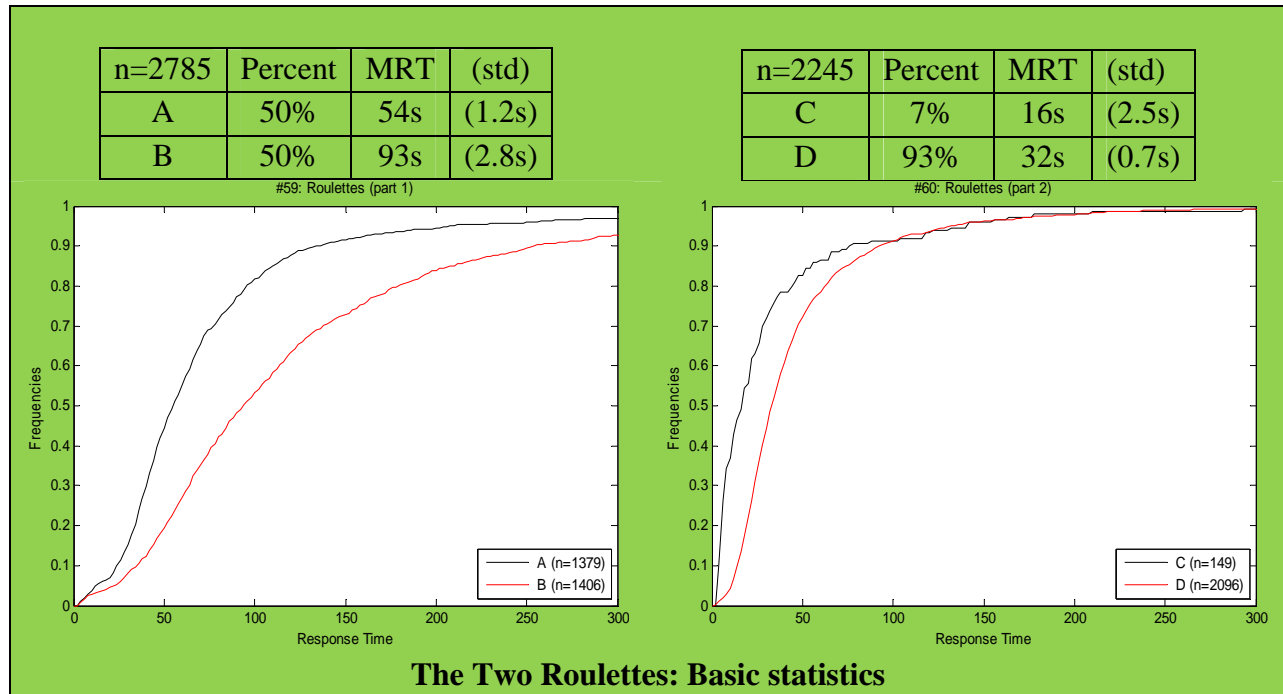
Among the local ranking fastest quartile, 70% chose A and 30% chose B while among the slowest quartile, 25% chose A and 75% chose B. The global ranking is a strong predictor of response time in this problem. While 61% of the fastest responders chose A, 64% of the slowest responders chose B.

	Local ranking quartiles				Global ranking quartiles			
Response	Fastest	Fast	Slow	Slowest	Fastest	Fast	Slow	Slowest
A	70%	59%	44%	25%	61%	56%	48%	36%
B	30%	41%	56%	75%	39%	44%	52%	64%
n	696	696	696	697	525	570	632	592
MRT	30s	56s	89s	185s	35s	59s	82s	142s

The Two Roulettes: Results according to local and global rankings

D. The Wason Experiment

Another interesting example is the Wason (1960) experiment:

> Suppose that there are four cards in front of you, each with a number on one side and a letter on the other. The cards before you show the following:
> 4 U 3 M
> Which cards should you turn over in order to determine the truth of the following proposition: If a card has a vowel on one side, then it has an even number on the other?

The right answer, i.e. U+3, is selected by only 10% of the subjects (n=2000, m=1773). The three most commonly made mistakes are 4+U (21%), all (23%) and U (13%). Results reveal a strong correlation between choosing the right answer and high response time.

n=2000	Percent	MRT
U	13%	75s
4+U	21%	88s
U+3	10%	124s
All	23%	92s

The Wason Experiment: Basic statistics
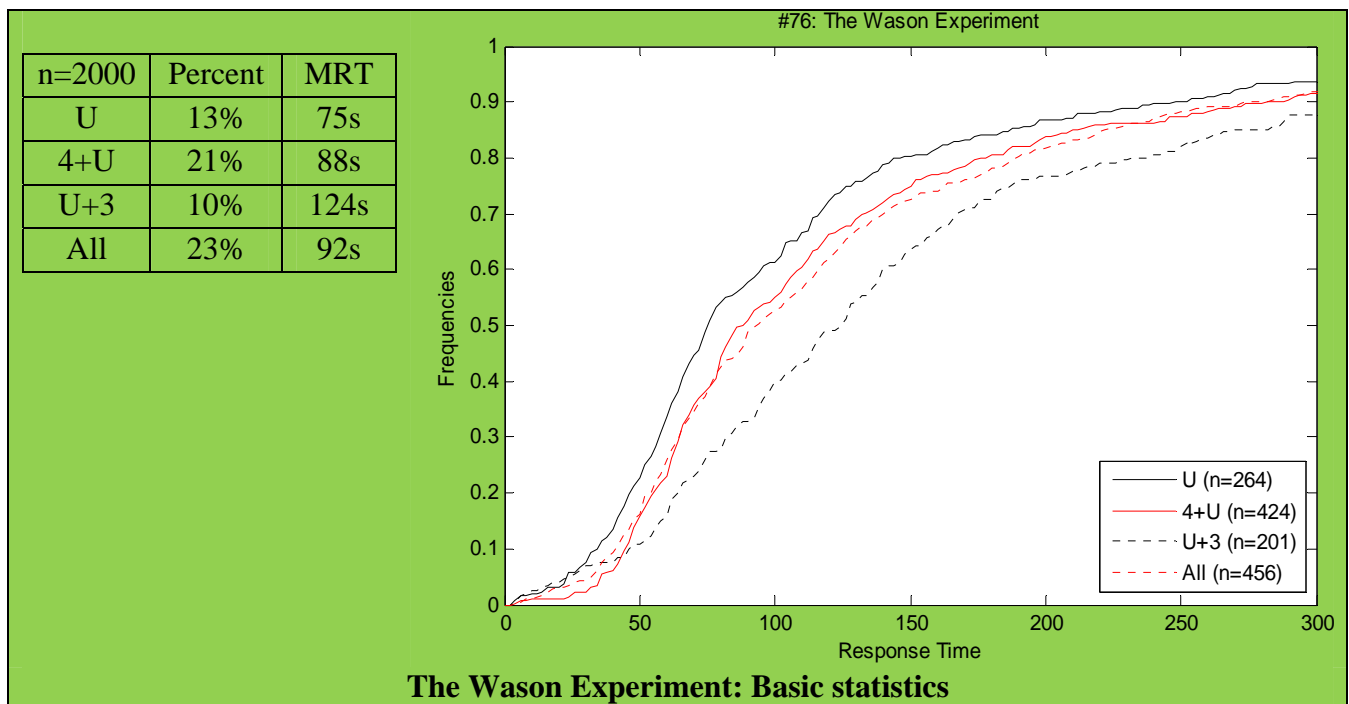
The low proportion of subjects who made the correct choice makes it difficult to draw further conclusions. What can be said is that according to the local ranking, 14% of the slowest quartile (MRT of 250 seconds) responded correctly in contrast to only 6% of the fastest quartile (MRT of 31 seconds). Similarly, there is a significant difference between the proportion of correct answers among the two extreme quartiles of responders as partitioned by the global ranking.

	Local ranking quartiles					Global ranking quartiles			
Response	Fastest	Fast	Slow	Slowest		Fastest	Fast	Slow	Slowest
U	15%	15%	12%	10%		13%	14%	11%	10%
4+U	17%	26%	21%	20%		20%	22%	22%	19%
U+3	6%	8%	12%	14%		7%	11%	10%	12%
All	20%	24%	23%	24%		18%	22%	29%	28%
n	500	500	500	500		480	466	420	407
MRT	38s	71s	116s	245s		47s	78s	114s	190s

The Wason Experiment: Results according to local and global rankings

3. Transitivity of Preferences and Response Time

The previous section presented four distinct problems, in which the notion of a mistake was well defined, to demonstrate the strong correlation between short response time and mistakes. This section presents the results of an experiment in which the correlation between mistakes and response time appears to be reversed and offers an explanation of the result.

Students in microeconomics courses (mostly on the PhD or MA level) answered a questionnaire consisting of 36 questions about nine vacation packages. In each of the questions, the subject was asked to compare between two packages (with three possible responses: 1. I prefer the first alternative; X. I am indifferent between the alternatives; and 2. I prefer the second alternative). Each vacation package was described in terms of four parameters: destination, price, level of accommodation and quality of food. The destination (either Paris or Rome) was always presented first, followed by the other three parameters in arbitrary order. Two of the alternatives were in fact identical:

   "A weekend in Paris at a 4-star hotel with a Zagat food rating of 17 for $574" and

   "A weekend in Paris for $574 with a Zagat food rating of 17 at a 4-star hotel".

   The goal of this rather tedious questionnaire was to demonstrate the concept of preferences and to demonstrate to the students that even they (i.e. economics students) often respond in a way that violates transitivity as soon as the alternatives become even slightly complicated. What constitutes a mistake in this case is clear.

   People are known to be embarrassed when they discover that their answers are not transitive (i.e. for some three alternatives, $x \succ y$, $y \succ z$ but not $x \succ z$, or $x \sim y$, $y \sim z$ but not $x \sim z$). When one of these two configurations appears in a subject's answers it is said that he exhibits a cycle.

15

The sample consists of 729 subjects who responded to this questionnaire. Only 12% of the subjects did not exhibit a violation of transitivity. The median number of cycles was 7 and the number of cycles varied widely -- from 0 to 58! Thus, even PhD students in economics often violate transitivity…

Distribution of Cycles
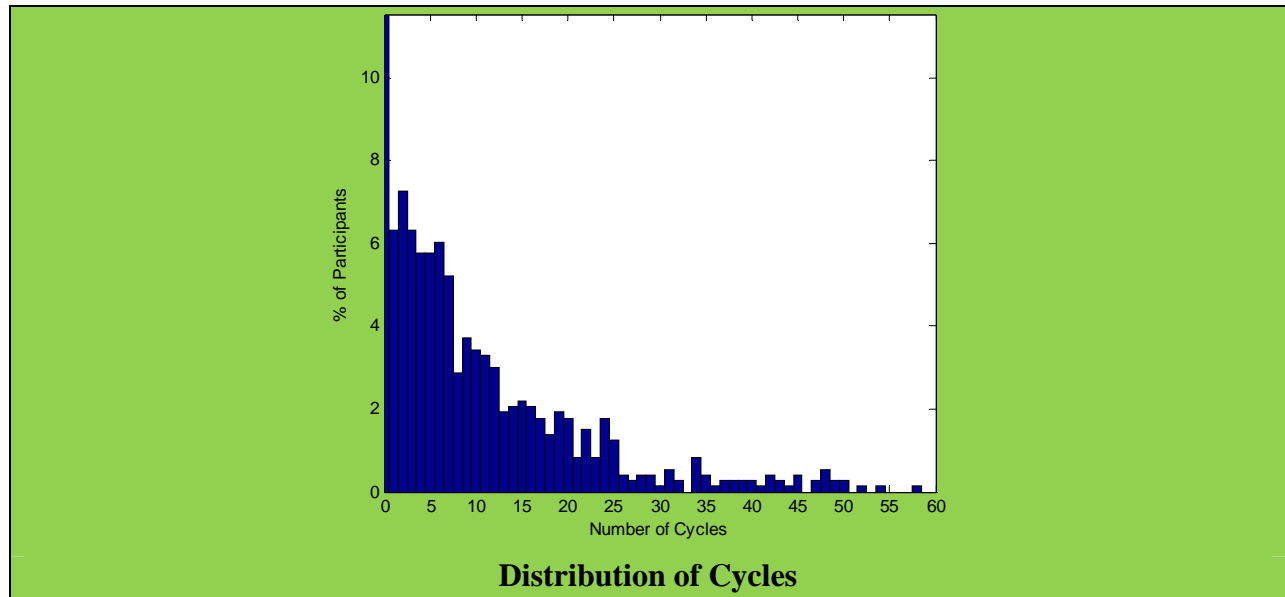
Interestingly, the most common cycle (29% of the total) involves the two alternatives mentioned above and "A weekend in Paris with a Zagat food rating of 20 at a 3-4 star hotel for $560".

The following diagram presents the cumulative response time distributions for subjects whose number of mistakes is below the median and above the median.

The Preference Questionnaire: RT distributions
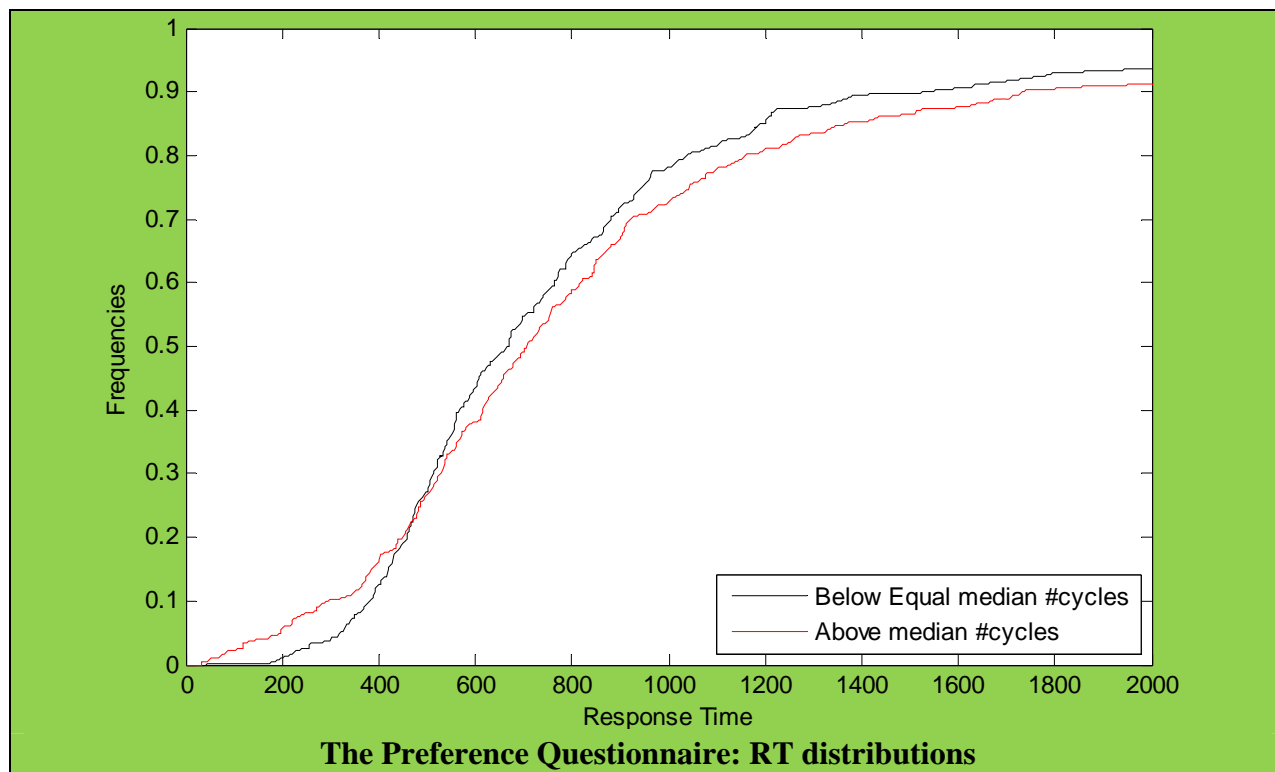
A comparison of the two graphs yields an unusual result. It appears that among the fastest quartile, there is a majority who answered the questionnaire without giving it sufficient attention, resulting in a large number of cycles. However, among the slowest three quartiles (who devoted more than 7 minutes to the questionnaire), a larger number of mistakes is associated with higher response time.

On the face of it, the results in this section conflict with those in the previous section. In contrast to Section 2, here we find an inverse correlation in which shorter response time is correlated with fewer mistakes. A plausible explanation is that consistent answers might be an outcome of activating a simple rule (such as minimizing the cost of the vacation) which does not require a long response time. Correctly applying more complicated rules which produce no intransitivities (such as "I prefer Paris to Rome, but in Paris I look for better food and in Rome for better accommodation) requires more time in order to maintain consistency and avoid mistakes. Thus, in this problem, having some inconsistencies might be an outcome of more sophistication (compare this point with the quite different conclusion of Choi, Kariv, Müller and Silverman (2011)).

This point is related to a more general methodological question regarding the practices of experimental economists. Experimenters often present subjects with a string of similar problems. Some do this in order to investigate the learning process but many do it in order to increase the number of observations. Since bringing subjects to the lab is expensive, they inflate

17

the number of observations by presenting each subject with multiple problems. As a result, after a few repetitions, subjects are likely to develop a rule of behavior. The behavior observed is then essentially the result of a single decision rule and each subject should be counted as only one observation.

4. Test cases: Allais, Ellsberg and Kahneman-Tversky

In this section, we discuss the results of three problems that are often used in the literature to demonstrate the high incidence of behavior that conflicts with established theories. Each of the problems consists of a pair of questions. The standard theories predict a one-to-one correspondence between the answers to the two questions. The common experimental protocol requires that subjects be randomly divided between the two problems, which makes it possible to treat the two populations as identical. If the proportion of those who chose a particular alternative in one question differs from that in the other question, this is interpreted as being inconsistent with the theory.

A different protocol is used here. In each of the examples, subjects were asked to answer the two questions sequentially in the order in which they are presented here. The results are robust to this change in the protocol and inconsistent behavior remains very common, even though subjects could have easily detected the inconsistency.

a. The Allais paradox
The following is Kahneman and Tversky (1979)'s well-known version of the Allais paradox:

A1: You are to choose between the following two lotteries:

Lottery A which yields $4000 with probability 0.2 (and $0 otherwise).

Lottery B which yields $3000 with probability 0.25 (and $0 otherwise).

Which lottery would you choose?

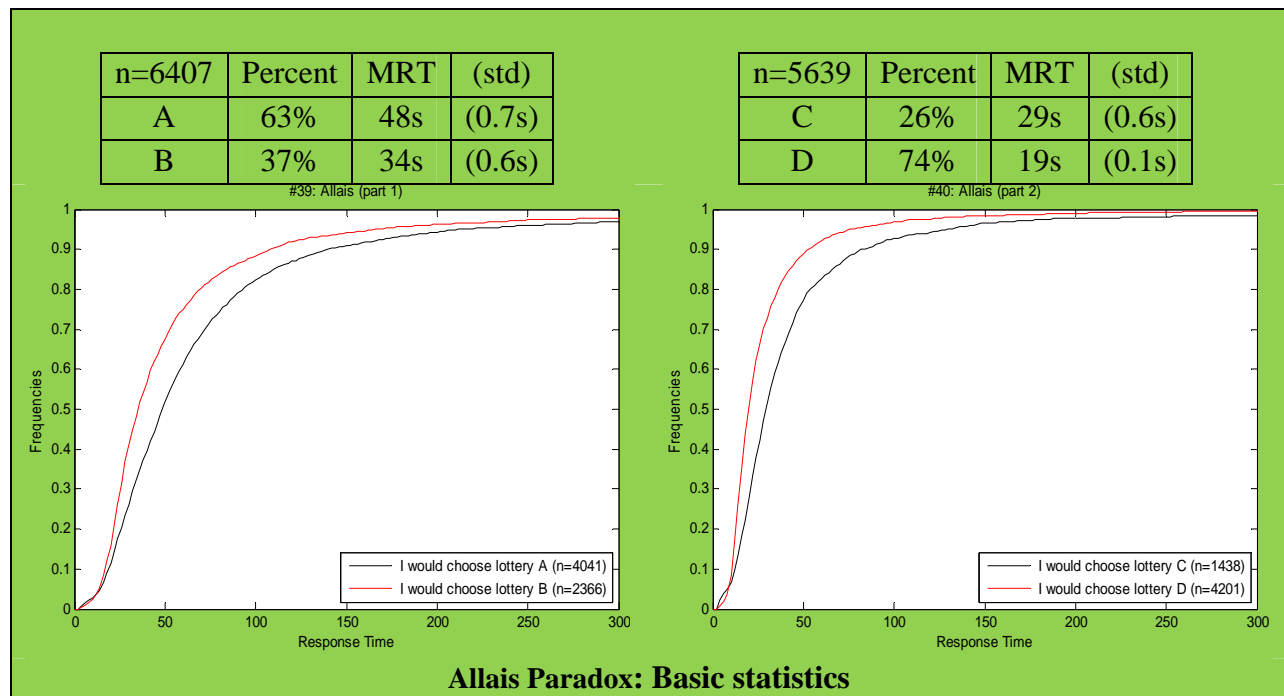A2: You are to choose between the following two lotteries:

Lottery C which yields $4000 with probability 0.8 (and $0 otherwise).

Lottery D which yields $ 3000 with probability 1.

Which lottery would you choose?

18

For completeness, we first examine the basic results (n=6407, m=5195 for A1 and n=5639, m=4588 for A2).

n=6407	Percent	MRT	(std)
A	63%	48s	(0.7s)
B	37%	34s	(0.6s)

#39: Allais (part 1)

n=5639	Percent	MRT	(std)
C	26%	29s	(0.6s)
D	74%	19s	(0.1s)

#40: Allais (part 2)

Allais Paradox: Basic statistics

Note that the choice of the less risky options (B and D) is associated with much shorter RT than the corresponding options (A and C, respectively). The fact that the response for A2 is faster than that for A1 is due to two factors: First, A2 was presented after A1 and therefore subjects were already familiar with this type of question. Second, based on the responses of those who answered only one of the two questions, it can be inferred that A2 is simpler than A1.

However, my main interest in analyzing these results lies elsewhere. Recall that the Independence axiom requires that the choices of a decision maker in each of the two problems be perfectly correlated. In other words, we should observe two types of decision makers: those who are less risk averse and choose A and C and those who are more risk averse and choose B and D. The standard typology of attitude towards risk (sometimes given by the coefficient of a CES utility function) specifies the characteristics that determine whether an economic agent chooses A and C or B and D.

The usefulness of RT in this case is in determining whether the behavior consistent with the Independence axiom is correlated with slow response time. There are 5528 subjects who responded to both problems, one after the other. The following table presents the joint distribution of their responses (the numbers in parentheses indicate the expected joint distribution if the answers to the two questions were completely independent):

19

n = 5528	C	D	Total
A	20% (16%)	44% (47%)	64%
B	5% (9%)	31% (27%)	36%
Total	25%	75%	100%

Allais Paradox: Joint Distribution of the Responses

Note that 49% of the subjects exhibited behavior that is inconsistent with the Independence axiom. The hypothesis that the answers to the two problems are independent is clearly rejected; however, the experimental joint distribution is not that far from the distribution expected if the answers to the two problems were totally independent. This finding (unrelated to RT) casts doubt on the basic typology that is commonly used to classify behavior under uncertainty. Would another typology perhaps explain the results better?

We now turn to the data on response time. A point on the following graph represents a single subject. The coordinates (x, y) indicate that the proportion x of the population responded faster than the subject did to A1 and the proportion y responded faster than he did to A2.
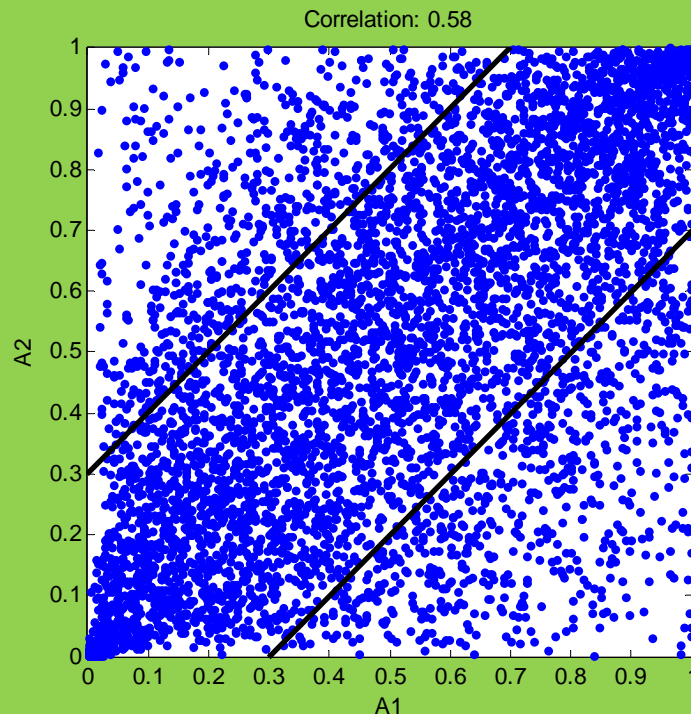
Allais Paradox: A graph of local rankings for A1 and A2

The correlation between the two relative positions of the points is quite high (0.58). Note that the vast majority of subjects fall within the area between the two diagonal lines. The high level of correlation suggests that we should examine the behavior of subjects who were consistently fast or slow in responding to this particular pair of problems.

The table below was created as follows: Subjects whose response time was above the median in both problems are classified as Slow and those whose response time was below the median in both problems are classified as Fast. 3966 subjects are included in the two categories. The table presents the two joint distributions of responses (fast subjects in the upper left-hand corner of the cell and slow subjects in the bottom right-hand corner):

Fast / Slow	C	D	Total
A	12% / 33%	43% / 42%	55% / 75%
B	5% / 5%	40% / 20%	45% / 25%
Total	17% / 37%	83% / 63%	n = 1983 / n = 1983

Allais Paradox: Joint Distribution according to Local Ranking

There are two striking patterns in the results: First, the choices of the slow group were no more consistent with the Independence axiom than those of the fast group. Thus, longer response time appears to contribute little to consistency. The choices of 52% of the fast subjects and 53% of the slow subjects were consistent with the Independence axiom. This result supports the view that the inconsistency of the choices with expected utility theory is not simply an outcome of error.

Second, the pattern of those choices that are consistent with expected utility theory differs dramatically between the groups. While almost 4/5 of the fastest consistent subjects chose B and D, less than 2/5 of the slowest consistent subjects chose that combination.

These results suggest a direction for new theoretical models to be based on a two-element typology (Rubinstein (2008)). There will be a Fast type who will either behave instinctively (choose A and D) or exhibit risk aversion (choose B and D), with equal probability. A Slow type will either behave instinctively (choose A and D) or maximize expected payoff (choose A and C), with equal probability. Further research is needed in order to establish such a typology.

b. The Ellsberg Paradox

The Ellsberg Paradox was presented to students in the following two versions (most students answered E1 first):

E1. Imagine an urn known to contain 30 red balls and 60 black and yellow balls (in an unknown proportion). One ball is to be drawn at random from the urn. The following actions are available to you:

"bet on red": yielding $100 if the drawn ball is red and $0 otherwise.

"bet on black": yielding $100 if the drawn ball is black and $0 otherwise.

Which action would you choose?

E2. Imagine an urn known to contain 30 red balls and 60 black and yellow balls (in an unknown proportion). One ball is to be drawn at random from the urn. The following actions are available to you:

"bet on red or yellow": yielding $100 if the drawn ball is red or yellow and $0 otherwise.

"bet on black or yellow": yielding $100 if the drawn ball is black or yellow and $0 otherwise.

Which action would you choose?

The results do not show any significant correlation between the responses of subjects to the two questions. The following table presents the joint distribution of their responses (once again the numbers in parentheses indicate the expected joint distribution if the answers to the two questions were completely independent):

		E2		
	n = 1791	red or yellow	black or yellow	Total
	Red	15% (13%)	49% (51%)	64%
E1	Black	6% (7%)	30% (28%)	36%
	Total	21%	79%	100%

Ellsberg: Joint Distribution of the Responses

However, my main interest is in the question of whether the "inconsistency" (in the form of choosing "red" in E1 and "black or yellow" in E2) is correlated with short response time. In the following table, Fast and Slow are defined as in the discussion of the Allais paradox, i.e. Fast subjects had response times below the median in both E1 and E2 and Slow subjects had response times above the median.

22

Ellsberg Paradox: Joint Distribution							
Fast / Slow	red or yellow	black or yellow	Total	Fast / Slow	red or yellow	black or yellow	Total
red	19% / 11%	46% / 50%	66% / 61%	red	18% / 11%	47% / 52%	65% / 63%
black	7% / 6%	28% / 33%	34% / 39%	black	6% / 6%	29% / 31%	35% / 37%
Total	26% / 17%	74% / 83%	n = 625 / n = 626	Total	24% / 18%	76% / 82%	n = 771 / n = 826
according to Local Ranking				**according to Global Ranking**			

The proportion of consistent subjects is almost the same in both groups (47% among the fast and 44% among the slow). Among the consistent subjects, there is a difference in behavior between the fast and slow groups, although less dramatic than in the Allais Paradox case. Similar joint distributions are obtained if we divide the subjects according to global ranking.

c. Framing (Kahneman and Tversky)

Kahneman and Tversky (1986) proposed the famous "outbreak of disease" experiment:

KT1. The outbreak of a particular disease will cause 600 deaths in the US. Two mutually exclusive prevention programs will yield the following results:

A: 400 people will die.
B: With probability 1/3, 0 people will die and with probability 2/3, 600 people will die.

You are in charge of choosing between the two programs.
Which would you choose?

KT2. The outbreak of a particular disease will cause 600 deaths in the US. Two mutually exclusive prevention programs will yield the following results:
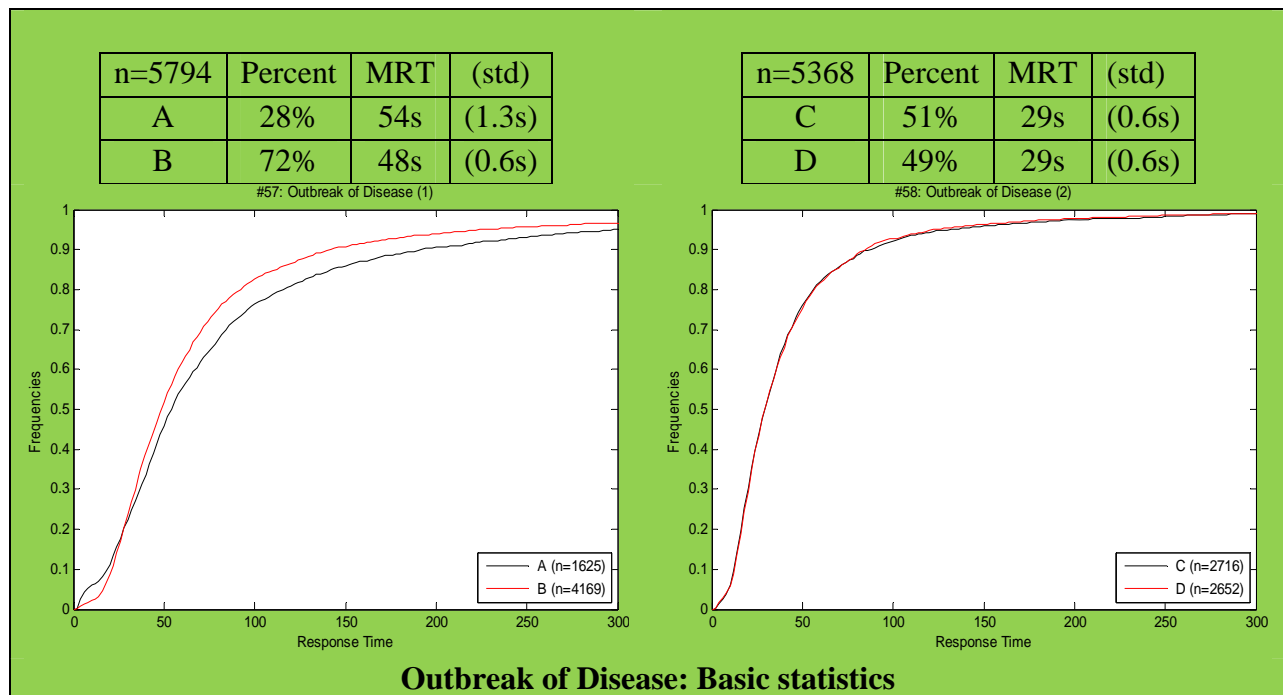
C: 200 people will be saved
D: With probability 1/3, all 600 will be saved and with probability 2/3 none will be saved.

You are in charge of choosing between the two programs.
Which would you choose?

The results are similar to those of Kahneman and Tversky. KT1 was presented first and 28% of the subjects (vs. 22% in the original experiment) chose A (n=5694, m=5221). B appears to be the more instinctive choice in this case (especially if we ignore the very fast-responding tail of the distribution). Remarkably, the two RT distributions for those who chose C or D in KT2 are almost identical. The fact that 49% of the subjects (vs. 28% in the original experiment) chose D in KT2 is partially a result of KT2 being presented to the same subjects subsequent to KT1 (Kahneman and Tversky followed the standard protocol and gave the two questions to two different populations). Note, however, that 44% of a small sample of 91 subjects who answered only KT2 chose D.

n=5794	Percent	MRT	(std)
A	28%	54s	(1.3s)
B	72%	48s	(0.6s)

#57: Outbreak of Disease (1)

n=5368	Percent	MRT	(std)
C	51%	29s	(0.6s)
D	49%	29s	(0.6s)

#58: Outbreak of Disease (2)

Outbreak of Disease: Basic statistics

n = 5277	C	D	Total
A	23% (14%)	5% (14%)	28%
B	28% (36%)	44% (35%)	72%
Total	51%	49%	100%

Outbreak of Disease: Joint Distribution of the Responses

A large proportion of subjects (67%) exhibited consistent behavior in this problem, whereas only 49% would be expected to do so if the choices were independent.

Outbreak of Disease: Joint Distribution							
Fast / Slow	C	D	Total	Fast / Slow	C	D	Total
A	22% / 24%	7% / 5%	29% / 29%	A	21% / 24%	6% / 5%	27% / 28%
B	27% / 28%	44% / 44%	71% / 71%	B	29% / 27%	44% / 45%	73% / 72%
Total	49% / 52%	51% / 48%	n = 1837 / n = 1838	Total	50% / 51%	50% / 49%	n = 2412 / n = 2409
according to Local Ranking				**according to Global Ranking**			

The distributions of the Fast and Slow halves, as classified by the local or global ranking, are remarkably similar. This tends to indicate that a typology of subjects in this context should not in fact be based on the Fast/Slow categories.

5. Other choice problems

This section examines two additional choice problems, each involving an ethical dilemma. It is included in order to emphasize that although response time might be an interesting indicator in some problems, it tells us very little about the nature of decision making in situations which clearly involve a moral dilemma.

A. The Kiosk Dilemma

Imagine you are the owner of a kiosk selling freshly-squeezed orange juice. You charge $4 per glass. A person arrives at the kiosk who is not a regular customer. He asks for a glass of diluted orange juice with equal amounts of juice and water and gives you 4 one-dollar bills. Which of the following statements would best describe your reaction?
  I would accept the $4.
  I would accept only $3 and give him back $1.

This dilemma is based on my own personal experience: I am used to asking for a diluted coffee (consisting of ¼ coffee and ¾ water). The vendors often feel that they should charge me only $1.25 rather than $1.50 (see my article http://arielrubinstein.tau.ac.il/articles/midnightEast.pdf).

In a sample of n=1130 (m=785), a majority (58%) chose to reduce the price. However, the cumulative RT distributions for the two choices are almost identical.

n=1130	Percent	MRT
Take $4	42%	50s
Give back $1	58%	50s

Kiosk: Basic statistics
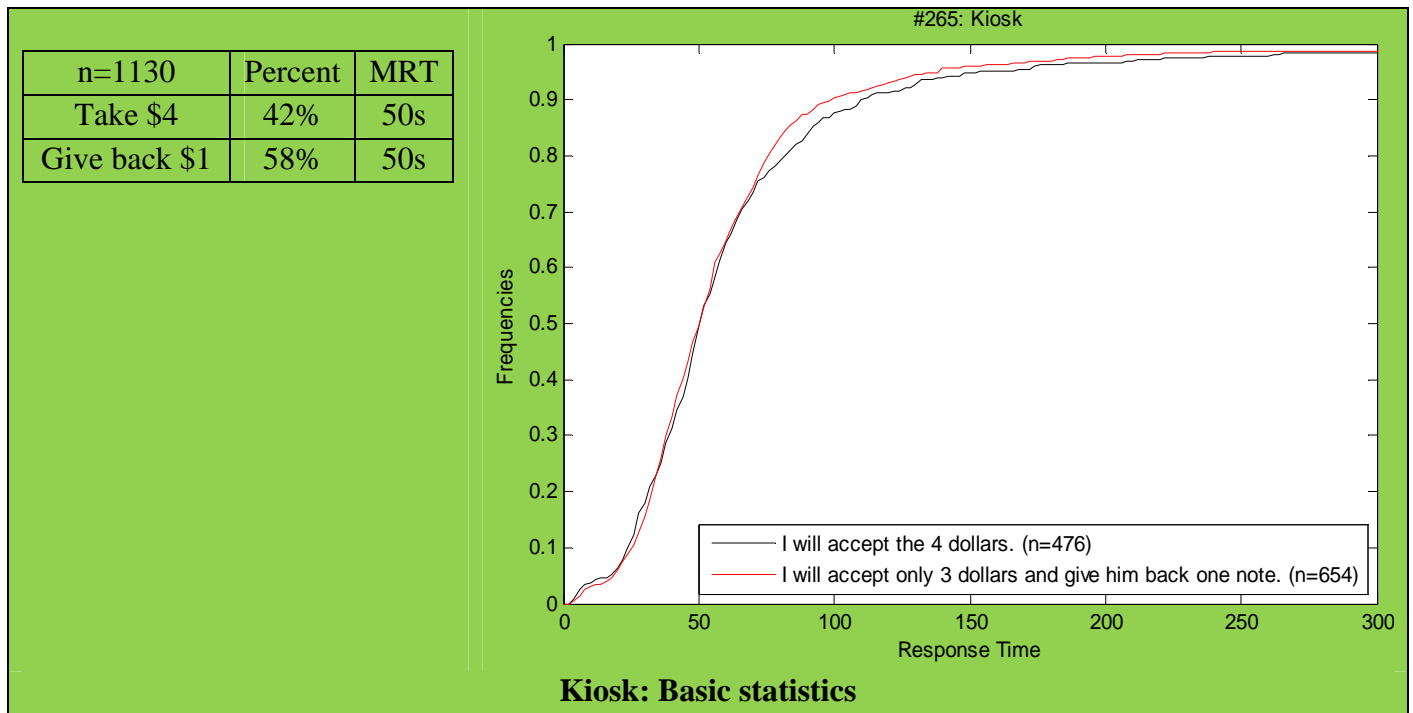
B. Manipulative Elections

> You are a member of a political party that is electing a new leader.
> There are three candidates: A, B, and C. According to the latest polls, the proportion of voters supporting each of the candidates is as follows:
>
>   Candidate A 44%
>   Candidate B 38%
>   Candidate C 18%
>
> Your favorite candidate is C and like all the other supporters of C, you prefer candidate B to candidate A.
> Candidate B is asking all C supporters to vote for him in order to help him beat candidate A.
> Which candidate would you vote for?

Only 4% of subjects (n=3826, m=3212) voted for A and they were particularly fast. There is only a small difference in the RT between the 76% who voted for B and the 21% who voted for C (the response time of the latter voters is somewhat longer). Voting for the first best is not the instinctive action and there is some evidence that voting manipulatively is more instinctive than voting for the favorite losing candidate.

26

N=3826	Percent	MRT
A	4%	35s
B	76%	55s
C	21%	59s

Manipulative Elections: Basic statistics
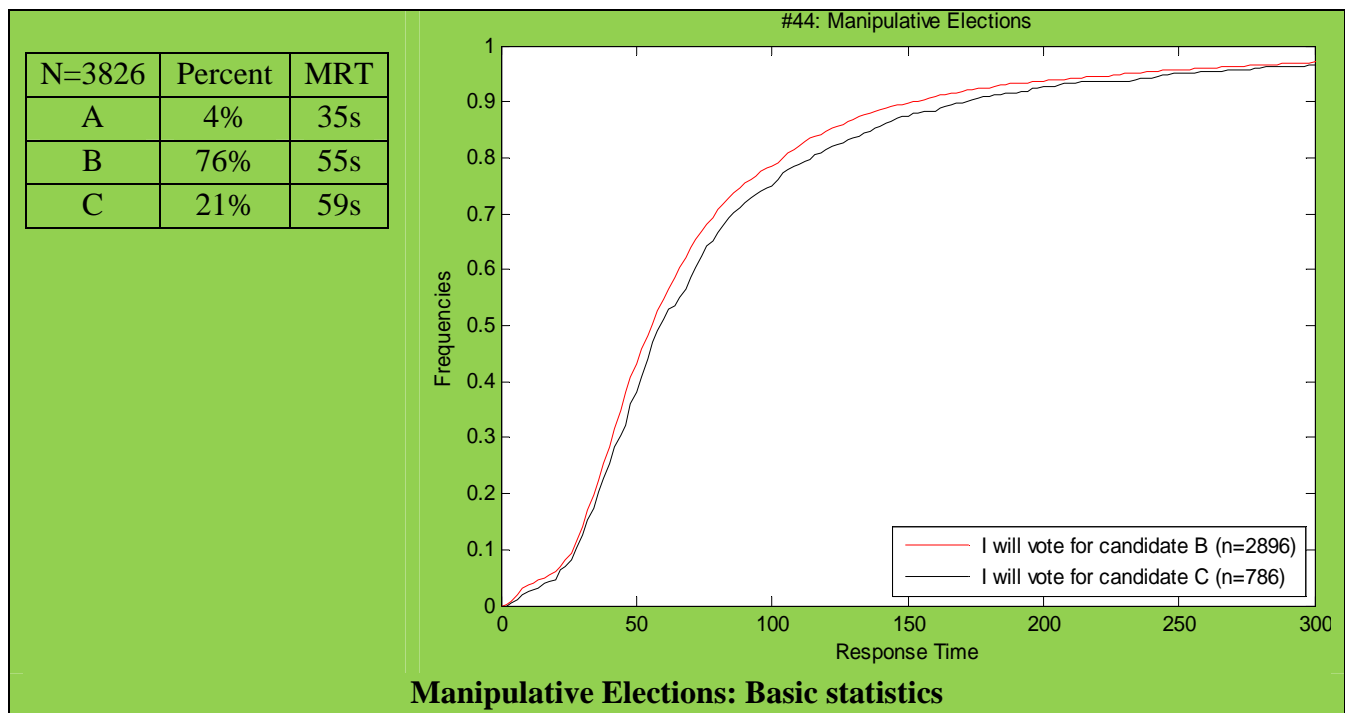
6. Final Comments

The line of argument presented above can be summarized as follows:

a) Section 2 presented four problems (counting the F's, comparing the likelihood of two sequences, choosing between roulettes and the Wason Experiment) in order to demonstrate that when the notion of a mistake is a clear cut there is a strong correlation between short response time and mistakes.

b) Section 3 showed that an inverse relation between short response time and making a mistake may appear in certain situations (such as when making 36 comparisons between pairs of alternative) due to the use of a simple choice rule that may be associated with making less mistakes.

c) Section 4 showed that short response time is not associated with behavior that conflicts with the standard theories, which are typically violated by Allais, Ellsberg and Kahneman-Tversky framing paradoxes. Therefore, behavior that is not consistent with Expected Utility Theory and is sensitive to framing effects is not the outcome of mistaken behavior but rather of more prolonged deliberation, which does not violate the approaches involving full rationality.

d) In some cases (especially that of the Allais paradox), the results hint at the usefulness of classifying subjects as fast or slow responders.

e) Section 5 presents two problems (fair price and speculative voting) to demonstrate that when a moral dilemma is involved response time is not correlated with choice.

I have often heard the claim that clever individuals will respond more quickly. Although we may see a number of people (especially in academia) who are both clever and respond quickly, the data presented in section 2 indicates that their number is small. Thus, for example, the small proportion of subjects who responded correctly to the Wason experiment (10%) had an RT that was about 30 seconds longer than the rest of the subjects. This result indicates that being both fast and clever is not a particularly common combination.

I began this research with the intention of looking for strong correlations between the choice behavior of subjects in different contexts. However, the data indicate only a weak correlation between the behavior of subjects in different problems, unless the problems are similar to each other. There are those who will be disappointed by such a finding. In contrast, I feel that the difficulty in predicting a person's behavior (based on his past behavior) is part and parcel of an individual's freedom of action.

Whatever reservations one might have regarding the above analysis, I hope that it has at least suggested that response time is an interesting tool in the evaluation of experimental results.

References

Arad, Ayala, and Ariel Rubinstein. 2012. "Multi-Dimensional Iterative Reasoning in Action: The Case of the Colonel Blotto Game." *Journal of Economic Behavior & Organization* 84 (2): 571-85.

Branas-Garza, Pablo, Debrah Meloso, and Luis Miller. 2012. "Interactive and moral reasoning: A comparative study of response times." *No.w440 Innocenzo Gasparini Institute for Economic Research*.

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman. 2011. "Who is (more) Rational?." *No. w16791. National Bureau of Economic Research*.

Christopher F. Chabris, David Laibson, Carrie L. Morris, Jonathon P. Schuldt, Dmitry Taubinsky. 2008. "Measuring intertemporal preferences using response times." *No. w14353. National Bureau of Economic Research*.

Donders, Franciscus Cornelis. 1969. "On the speed of mental processes." *Acta psychologica* 30: 412-31.

Kahneman, Daniel and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47: 263-92.

Kahneman, Daniel. 2011. "Thinking, fast and slow." *Farrar, Straus and Giroux.*

Lotito, Gianna, Matteo Migheli, and Guido Ortona. 2011. "Is cooperation instinctive? Evidence from the response times in a public goods game." *Journal of Bioeconomics* 13: 1-11.

Piovesan, Marco, and Erik Wengström. 2009. "Fast or fair? A study of response times." *Economics Letters* 105: 193-6.

Rubinstein, Ariel. 2001. "A Theorist's View of Experiments." *European Economic Review* 45: 615-28.

Rubinstein, Ariel. 2007. "Instinctive and Cognitive Reasoning: A Study of Response Times." *The Economic Journal* 117: 1243-59.

Rubinstein, Ariel. 2008. "Comments on NeuroEconomics." *Economics and Philosophy* 24: 485-94.

Schotter, Andrew, and Isabel Trevino. 2012. "Is response time predictive of choice? An experimental study of threshold strategies." Unpublished.

Sternberg, Saul. 1969. "Memory-scanning: Mental processes revealed by reaction-time experiments." *American scientist,* 57: 421-57.

Studer, Raphael, and Rainer Winkelmann. 2012. "Reported happiness, fast and slow." Unpublished.

Tversky, Amos and Daniel Kahneman. 1983. "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment." *Psychological Review* 90 (4): 293-315.

Tversky, Amos and Daniel Kahneman. 1986. "Rational Choice and the Framing of Decisions.**"** *The Journal of Business* 59 (4): 251-78.

Wason, Peter. C. 1968. "Reasoning about a rule." *The Quarterly Journal of Experimental Psychology* 20 (3): 273–81.