Keldenich, Klemens; Klemm, Marcus

Working Paper

Double or Nothing!? Small Groups Making Decisions Under Risk in ""Quiz Taxi""

Ruhr Economic Papers, No. 278

Provided in Cooperation with:
RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

Suggested Citation: Keldenich, Klemens; Klemm, Marcus (2011) : Double or Nothing!? Small Groups Making Decisions Under Risk in ""Quiz Taxi"", Ruhr Economic Papers, No. 278, ISBN 978-3-86788-323-8, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Essen

This Version is available at:
https://hdl.handle.net/10419/61462

RUHR
ECONOMIC PAPERS

Klemens Keldenich
Marcus Klemm

Double or Nothing!? Small Groups Making Decisions Under Risk in "Quiz Taxi"

#278

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, email: W.Leininger@wiso.uni-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Christoph M. Schmidt
RWI, Phone: +49 (0) 201/81 49-227, e-mail: christoph.schmidt@rwi-essen.de

Editorial Office

Joachim Schmidt
RWI, Phone: +49 (0) 201/81 49-292, e-mail: joachim.schmidt@rwi-essen.de

Ruhr Economic Papers #278

Responsible Editor: Thomas K. Bauer

ISSN 1864-4872 (online) – ISBN 978-3-86788-323-8

Ruhr Economic Papers #278

Klemens Keldenich and Marcus Klemm

Double or Nothing!? Small Groups Making Decisions Under Risk in "Quiz Taxi"

RUHR UNIVERSITÄT BOCHUM RUB

UNIVERSITÄT DUISBURG ESSEN

Ruhr Graduate School in Economics

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

ISSN 1864-4872 (online)
ISBN 978-3-86788-323-8

Klemens Keldenich and Marcus Klemm[1]

Double or Nothing!? Small Groups Making Decisions Under Risk in "Quiz Taxi"

Abstract

This paper investigates the behavior of contestants in the game show "Quiz Taxi" when faced with the decision whether to bet the winnings they have acquired on a final "double or nothing" question. The decision is made by groups of two or three persons. This set-up enables the decision making process to be studied by observing group communication. There is a strong correlation between communication content and the final choice, indicating that, from the contestant's perspective, the decisions are rational and that the context is an important factor in the final decision. This is particularly so for individual valuations of the money at stake. More extensive discussions help to make the right decision. As contestants do not apply to go on the show, they represent a less selected sample than those in previous game show studies. Overall, the contestants show risk averse behavior, suggesting CRRA-parameters larger than 1. The study also shows some heterogeneity in attitude to risk. Contestants who do better in the show are more likely to go for the risky option, because they are more knowledgeable and more confident. All-female groups are less likely and three-person groups more likely to choose the risky option.

JEL Classification: C93, D70, D81

Keywords: Risk attitude; game show; communication; group decision

August 2011

---

1 Klemens, Keldenich, RGS Econ, UDE; Marcus Klemm, RGS, RUB. *– We thank Jeannette Brosig-Koch, John Haisken-DeNew, Thomas Bauer, as well as seminar participants at the 4th RGS Doctoral Conference in Economics 2011, the 2011 Annual Conference of the Scottish Economic Society, the 1st RGS Jamboree 2010, the University of Duisburg-Essen, and the Ruhr-University Bochum for helpful comments and suggestions. We also thank two "Quiz Taxi" contestants for providing us with additional information, and Michael Klemm for help with the tool that was used to record the data. Financial support by the RGS Econ is gratefully acknowledged. Marcus Klemm also thanks the Ruhr-University Research School funded by Germany's Excellence Initiative for further support [DFG GSC 98/1].* *– All correspondence to Klemens Keldenich, RGS Econ, c/o Chair for Economic Policy, University of Duisburg-Essen, Faculty of Economics, Universitätsstr. 12, 45117 Essen, Germany, E-Mail: klemens.keldenich@ibes.uni-due.de.*

1 Introduction

Risk attitudes are an integral part of many economic decisions and thus economic models. However, the empirical assessment of risk attitudes has proven to be very challenging because many confounding factors make isolating risk attitudes difficult. Previous economic studies have used several methods, including estimation of life-cycle macro-models, self-reported measures from micro-surveys, and laboratory experiments, to quantify risk attitudes. Each of these approaches has its shortcomings which include, e.g., the failure to account for unobserved individual heterogeneity, sample selection, or lack of incentives to reveal true preferences.

Since Gertner (1993) and Metrick (1995), economic research has been using data from TV game shows which is generated in a way that lends itself to measuring risk attitudes. This paper adds to this strand of research on decision making under risk by analyzing the behavior of contestants in the German TV game show "Quiz Taxi".[1] The "Quiz Taxi" is unique among game shows because contestants do not apply actively to be on the show, which should reduce any potential selection bias. They participate in a quiz while driving in a cab. Having reached the destination, the candidates are asked if they want to play a final "master question". If they answer correctly, their winnings will double; if they are wrong, they will lose all.

The paper makes two main contributions: firstly, the group decision about this master question is used to study risk attitudes, focusing on possible sources of heterogeneity pertaining to sex, age, and group size. Secondly, the data also allows a study of group decision making by analyzing the communication which leads to the final decision. The agents in this particular setup are groups that must reach a consensus. The individuals must therefore communicate their preferences, which in most other settings remain unobservable to researchers. Here, the "black box" of decision making can be opened to some extent. As many economic decisions are made by groups rather than individuals[2] and there might be a systematic difference in behavior between groups and individuals, this is an important aspect of the data.

In line with conventional theory, the empirical analysis reveals that contestants are risk averse. Despite fairly high chances of answering the master question correctly, only one third of all groups decides to play the master question. The tendency to take the risky master question decreases with the winnings at stake and increases with the actual performance during the cab ride. With regard to observable characteristics, all-female groups appear to be more risk averse and three-person groups more inclined to take the risk. The communication analysis reveals very strong correlations between the communication content and the decision made. This suggests that contestants make rational decisions from their subjective point of view.

[1] The game show "Quiz Taxi" is the German version of the English show "Cash Cab" which first aired in 2005. It was sold to over 25 countries and is still running in several of those, e.g., USA and Australia. Bliss et al. (2011) infer coefficients of absolute risk aversion from the behavior of contestants in the US version of the show.

[2] Examples include household decisions about consumption and savings, "virtually all significant strategic decisions by corporations" (Cooper and Kagel 2005), central bank decisions regarding monetary policy (Blinder 2007), or investment decisions by mutual funds (Prather and Middleton 2002).

While about half of the groups immediately agree not to play, more than half of the initially undecided groups decide to play the risky master question. The initial response therefore appears to be dictated by risk aversion, which decreases with the length of the discussion and the arguments exchanged between group members.

The remainder of the paper is organized as follows. The next section briefly discusses the empirical background of the assessment of risk attitudes, focusing on research based on TV game shows. The "Quiz Taxi" and its contestants are described in section 3. The empirical analysis and results are presented in section 4 and the conclusions in section 5.

2 Empirical background

A broad range of literature about the empirical assessment of risk attitudes has developed, using several different sources of data. The various approaches are plagued by different methodological problems. Empirical macroeconomic studies try to infer risk attitudes by estimating life-cycle models of consumption. Typically, such studies assume constant relative risk aversion (CRRA) utility functions and estimate the Arrow-Pratt measure of relative risk aversion. Two early examples are Hall (1988) and Attanasio and Weber (1989) who both find fairly high degrees of risk aversion. While such estimates are useful as a description of average behavior (e.g., for the parameterization of macroeconomic models), they do not say much about individual decisions made under risk. In particular, risk aversion cannot be disentangled from other preferences, e.g., for intertemporal substitution. Furthermore, a considerable degree of heterogeneity in risk attitudes has been found depending on gender (Jianakoplos and Bernasek 1998, Bliss et al. 2011), group size (Baker II et al. 2008), or age (Donkers et al. 2001).

For these reasons, microeconomic data is needed to study decision making under risk or uncertainty.[3] Surveys often include questions on risk attitude, either in the form of hypothetical gambles and lotteries, or as direct self-assessments. Problems with using this kind of data include the subjective nature of the answers and the lack of incentives to answer truthfully. Dohmen et al. (2009b) relate such data from the German Socio-Economic Panel (SOEP) to experiments with the same SOEP respondents and find that the general risk question in the survey is a valid measure of risk attitude. Their analysis suggests that most individual CRRA parameters lie between 1 and 10. They also show that risk attitude varies with age, gender, and parental background. Barsky (1997) finds evidence for relatively high relative risk aversion (CRRA parameter of 12) with a very high variance using data from a hypothetical decision over lifetime income from the US Health and Retirement Survey. In a similar approach, Donkers et al. (2001) use the CentER Savings Survey with Dutch households to estimate risk attitudes. Their semi-parametric estimates reject the assumptions of expected utility theory. Instead, decisions are better described using

[3] Since "Quiz Taxi" contestants know about the design of the decision problem and must have subjective probabilities for the possible outcomes, their decisions are made under risk, not under uncertainty. Therefore, the term "risk" is used in the following (see Knight 1921, for the seminal contribution about the difference between risk and uncertainty).

prospect theory (Kahneman and Tversky 1979). It is found that decisions (and thus risk attitudes) vary with age, income, and individual wealth.

The most natural approach to studying risk attitudes are probably (laboratory) experiments where subjects must make decisions under risk with real monetary consequences in controlled environments. This method alleviates the problems of subjectivity and lack of incentives but possible problems remain, including small monetary incentives, sample selection, and experimenter effects, which would all lead to a low external validity. The seminal paper in this area by Holt and Laury (2002) compares gambles over different stakes, and real gambles with hypothetical ones. They find that a significant difference between real and hypothetical gambles emerges for larger gambles, and that the coefficient of relative risk aversion increases with the stake (from 0.3 to 0.9). Baker II et al. (2008) use an experiment to look at differences between groups and individuals and find an interaction effect: When the gamble contains either very high or very low probabilities, groups are less risk averse than individuals, while there is no difference in gambles with medium probabilities. The findings of Rockenbach et al. (2007) suggest that groups are more efficient, and better at making the trade-off between risk and higher expected payoff.

TV game shows have become a popular and valuable source of data for the study of decision making under risk (for a survey, see Andersen et al. 2008). They offer clearly defined decision situations and parameters, including the decision space, the decision maker, and the stakes. Therefore, game shows can be regarded as natural experiments (see List 2006) because conditions between subjects are kept constant, enabling the effect of one variable to be studied with less interference than with survey data. Furthermore, the stakes are usually substantial. Due to these positive attributes, Harrison and List (2004) call situations which generate such data a "serendipity". This property was first exploited by Gertner (1993) who looks at the UK game show "Card Sharks". In this show, contestants are asked to wager a part of their winnings in a situation with clear probabilities over the outcomes. The coefficient of relative risk aversion is estimated at 4.8, but behavior is found to be incompatible with expected utility maximization. In contrast, Metrick (1995) finds that players in the show "Jeopardy!" behave in an almost risk-neutral way. Hersch and McDougall (1997) also find evidence for risk-neutrality of contestants in the lottery game show "Illinois Instant Riches" which involved very high stakes from USD -10,000 to USD 15,000. Using data from another lottery game show "Hoosier Millionaire"with high stakes of up to USD 1,000,000, Fullenkamp et al. (2003) find that contestants are risk-averse, but that the degree of risk aversion varies with the stakes.

More recently, Post et al. (2008) look at the US, Dutch and German versions of the show "Deal or No Deal".[4] Similar to the study by Gertner (1993), they find behavior to be incompatible with standard expected utility theory. Instead, the data is best organized using alternatives which are path-dependent, such as prospect theory. They can explain large parts of behavior with outcomes experienced earlier in

[4]"Deal or No Deal" has become very popular among economists. See De Roos and Sarafidis (2009) and Gee (2007) for further investigations.

the game, even though the stakes are high and the decision situation is clearly defined. Brooks et al. (2009) show that the risk aversion of Australian "Deal or No Deal"-contestants increases with the stake, and varies with gender and age, but not with wealth.

Matsen and Strøm (2010) also discard expected utility theory (but also most other competing theories) for the explanation of contestants' behavior in the Norwegian game show "Joker" which involves large stakes of more than EUR 30,000 on average.[5] Similar to the decision that the "Quiz Taxi" contestants have to make, the "Joker" contestants can stop and keep their acquired winnings, or continue to play a risky gamble. Surprisingly, only 1 of 356 players has ever stopped. Another study which is based on a setup similar to the one at hand is Beetsma and Schotman (2001) who use data from the Dutch game show "Lingo". "Lingo" contestants have to decide up to five times if they want to continue playing with the possible outcomes of doubling their winnings or losing everything, or if they want to stop playing and keep what they have won so far. Furthermore, the contestants are always two-person groups, not individuals. Beetsma and Schotman (2001) find a high degree of risk aversion in both CARA and CRRA specifications, but do not report anything about the interaction or communication between the contestants.

Hartley et al. (2006) look at the popular game show "Who wants to be a millionaire?". This study is close to the study presented here in the sense that contestants must answer knowledge questions. Hence, their subjective assessment of their own ability has to be taken into account. Hartley et al. (2006) find that a standard CRRA framework fits the data very well. The coefficient of relative risk aversion is estimated at close to unity.

Another aspect pertinent to assessing risk attitudes is the question if the decision maker is a group or an individual. In "Quiz Taxi", the decision maker is always a group, either with two or with three persons. This may have an influence on risk attitude: A study by Rockenbach et al. (2007) compares groups with individuals in a portfolio selection task and finds that groups are better at making the trade-off between risk and higher expected payoff. The "Quiz Taxi" data allows studying the communication in the group, making it possible to analyze what drives group behavior. In this respect, the study presented here goes significantly beyond the work of Bliss et al. (2011) that focuses exclusively on the estimation of coefficients of absolute risk aversion for different types of groups. They report that teams perform better and are more likely to go for the risky option. The group decision appears to be based rather on the overall money at stake than on the individual stakes of each contestant involved. However, Bliss et al. (2011) do not report anything about the subjective decision making process of the contestants.[6]

[5] They find evidence for systematic expectation biases and systematic calculation errors.

[6] The authors only note that "The players are given a few moments to discuss their decision." (Bliss et al. 2011, p.7), but do not incorporate this feature of the show in the analysis.

3 Data

3.1 The game show

German "Quiz Taxi" was shown from 2006 to 2008 on the TV station "kabel eins" with a total of more than 500 episodes with 2 to 5 cab rides each. For the analysis presented here, over 100 publicly available episodes from 3 DVD's and the internet portal "maxdome.de" are used.[7] Overall, the set-up of the German "Quiz Taxi" is very similar to that of the US "Cash Cab" as described by Bliss et al. (2011). Notable differences are the group sizes (1 to 5 in the US compared to 2 or 3 in the German version) and the group discussions (apparently rather limited in the US, more extensive in Germany), which explains the different focus of the study presented here and that by Bliss et al. (2011).[8]

Data from TV shows has of course some limitations (see List 2006). Most importantly, the contestants are likely to represent only a selected sample of the whole population. This is particularly true for shows that are recorded in front of a large audience and very popular (e.g., "Who wants to be a Millionaire?"), and for shows that select participants from a specified pool of candidates (e.g. holders of a lottery ticket as in "Deal or No Deal" and "Joker"). The selection of contestants for "Quiz Taxi" is less critical in this respect: firstly, contestants do not apply actively to be on the show, but rather are quite randomly selected on the street. In an interview, the host Thomas Hackenberg says that he waits at taxi stands, gets leads from local taxi dispatchers, and often just picks up people hailing a cab (the "Quiz Taxi" is disguised as a regular cab).[9] Some contestants are also recruited on the street under false pretenses, e.g., they are asked if they would participate in a social or a market research project involving a cab ride.[10]

Regardless of the way the potential contestants are recruited, almost all of them enter the "Quiz Taxi" unaware that they can participate in a game show. The real nature of the show is not revealed to them until they get into the cab. They can then decide whether they want to play or not which obviously produces some problems of selection. In the interview mentioned above, the host notes that the reasons for backing out of the cab are very diverse.[11] It is important to bear in mind that certain population groups, e.g., fairly confident, risk loving people, and more affluent people who are able to afford taking a cab, might be overrepresented by the contestants. However, judging from the appearance and behavior of the contestants, they do not appear to be a highly selected group. In any case, selection should be much less of a problem for "Quiz Taxi" than for any other game show studied before. If there is any bias

[7] For more information about the show, see http://www.kabeleins.de/doku_reportage/quiz_taxi and http://www.kabeleins.de/serien_shows/quiz_taxi/artikel/12845 (links as of February 28, 2011).

[8] Further minor differences relate to the money earned for correct answers, and special tasks like the "red light challenges".

[9] See http://www.kabeleins.de/doku_reportage/quiz_taxi/artikel/20708 (link as of February 28, 2010). Due to the show's increasing popularity, the host even disguises in later episodes to keep the nature of the cab hidden.

[10] This information stems from two participants. One team was offered EUR 20 each for driving in a cab operated by an unemployed person, and judging the driver's performance. After the ride, they were taken back to the starting point. The other team was neither offered any money nor taken back to the start.

[11] The host speculates that some people are afraid that they might make a fool of themselves, simply do not want to be on TV, or are just accompanied by the wrong persons. In two episodes, people who declined to play were shown. Their reasons for not participating were time constraints. Unfortunately, we do not have any concrete information about the share of participants who decline to take part.

at all, it should be in the direction of choosing the risky alternative. The analysis presented below can therefore be seen as conservative estimates of risk averse behavior.[12]

Secondly, the framing of this field experiment should also be much less critical than in other quiz shows. Contestants are not placed in a TV studio but in the more familiar surroundings of a cab. In addition, there is no studio audience but only the host of the show, and the show is neither aired on one of the five biggest German TV stations nor at prime time. Bliss et al. (2011) also note that contestants cannot prepare for this particular quiz in advance.

FIGURE 1: Screenshot of "Quiz Taxi"

Source: kabel eins, Mondo Entertainment: Quiz Taxi DVD 1-1 (2008)

Figure 1 provides a screen shot of two "Quiz Taxi" contestants. The contestants are always groups of two or three people playing together.[13] The quiz consists of general knowledge questions (culture, history, sports, etc.) which are posed while the contestants are driven to their destination. Answering the questions correctly yields EUR 50 for the first three questions, then EUR 100 and after the 10th question EUR 150.[14] Missing a question leads to the loss of one "life", losing three lives means that the contestants lose all the money they have earned so far and have to leave the cab even if they have not reached the destination yet. There are also two "wild cards": The contestants can call somebody or they can ask passersby for help.[15] If the contestants manage to reach their destination without losing all three lives, the host shows them the money they have earned so far and asks them if they would like to play the "master question". If they answer this final question correctly, their winnings are doubled, otherwise

[12] There is one additional layer to the selection problem. The producers might only air episodes that they think are interesting enough. This seems to be the case for the cab rides that were chosen for the DVD's, but not for those from "maxdome.de". The analysis therefore includes a dummy identifying 14 of 146 observations that come from the DVD's. If such a selection process were important, estimates would again be biased into the direction of risk-loving behavior. From a producer's point of view, this would also be highly cost inefficient.

[13] This constitutes a major difference compared to the US version where group sizes of 1 to 5 persons enable Bliss et al. (2011) to investigate differences in risk attitudes between individuals and small groups.

[14] At the beginning of the first season, the first 5 questions were worth EUR 50, and each of the following EUR 100. Since in these episodes no master question was offered, they are disregarded for the analysis presented here.

[15] In later episodes, contestants can win back the previously used passersby-wild card by successfully solving an additional question or task when the cab has to stop at a red traffic light.

they lose everything. For economic research, this is the most interesting part of the show because it enables the decision to choose this risky option instead of leaving the cab with the acquired money to be studied. To this end, observable characteristics of the cab ride, the contestants, and the questions are recorded: city, distance from start to destination, group size, a contestant's sex, age[16] and possibly migration background[17], number of questions, final stake, share of correct and known answers (no wild card used, no obvious guesses), lives lost, use of wild cards.[18]

The main innovation of this paper is that it does not stop at these observable characteristics. The group discussion is transcribed as soon as the host asks the contestants if they want to play the master question until the final decision is reached. This allows aspects of the communication to be incorporated into the analysis, including the discussion time and content, initial opinions and discussion shares of group members. Since the process of recording this data is subject to measurement error, all information obtained from the transcripts was cross-checked. Furthermore, controlling for the person who coded the data does not have any influence on the obtained estimation results presented in section 4.[19]

3.2 Descriptive statistics

In total, 367 "Quiz Taxi" rides are publicly available. After excluding all rides from the beginning of the first season which were played by different rules without a master question, a few duplicates and a few rides with only one passenger, 256 cab rides with 562 individuals are left.[20] The analysis takes place on the group level because the final decision must be made by the group, and because the focus lies on the communication leading to that decision. The individual characteristics are hence aggregated to the group dimension.

Of the 256 groups, 43% do not reach the master question, 37.5% reach the master question and choose not to answer it and the remaining 19.5% choose to answer the master question. Table 1 gives the descriptive statistics by the success of the groups. Those who do not reach the final destination have of course answered fewer questions correctly and thus earned less money (which they cannot keep anyway). With regard to group characteristics, all-female groups reach the master question less often. None of the other group characteristics (size, age composition, migration background) seem to make a difference.

Looking only at the groups who reach the master question, the average stake for those who play the master question is EUR 771 (with a standard deviation of EUR 222) while those who decline to play have earned on average EUR 833 (standard deviation of EUR 255). Thus, a sizable range of stakes exists to test for a possibly stake-dependent decision. The two groups do not appear to differ much with regard

[16] The age is estimated by appearance of the contestants if no direct information is revealed during the show. Three broad categories are used: below 30 years of age, 30 to 50 years of age, and above 50 years of age.
[17] Based on language skills and appearance.
[18] Table A in the appendix provides an overview of the variable definitions.
[19] This is done by adding an indicator variable for the person who recorded the data to the regressions. The results never change. The dummy variable itself is always insignificant.
[20] The data also includes one ride which was available on the "kabel eins" website for some time. There were 5 duplicates, and 5 rides with only 1 passenger (including one celebrity).

to all other characteristics, including the share of questions answered correctly or known.

TABLE 1: Variable means (standard deviations) by success of contestants

	All contestants	Master question not reached	Master question reached ... but not played	and played	and solved
Stake (/100)	6.85	5.17	8.33	7.71	7.87
	(2.74)	(2.14)	(2.55)	(2.22)	(2.34)
Stake per contestant (/100)	3.18	2.40	3.87	3.58	3.66
	(1.30)	(1.05)	(1.16)	(1.16)	(1.23)
Distance	4.90	5.14	4.80	4.57	4.67
	(1.43)	(1.54)	(1.37)	(1.20)	(1.24)
Number of questions	10.18	9.44	10.98	10.26	10.37
	(2.13)	(2.06)	(2.06)	(1.87)	(1.96)
Lives left	0.77	0.00	1.25	1.52	1.54
	(0.81)	(0.00)	(0.54)	(0.71)	(0.74)
Share of correct answers (in %)	76.72	66.55	83.87	85.35	85.61
	(11.28)	(7.91)	(5.54)	(7.38)	(7.83)
Share of known answers (in %)	59.48	51.65	64.71	66.67	66.70
	(14.23)	(14.04)	(10.60)	(12.41)	(12.31)
Streak of correct answers	1.55	0.00	2.55	3.04	3.20
	(2.44)	(0.00)	(2.38)	(3.21)	(3.41)
Wild card left	0.21	0.19	0.18	0.30	0.29
	(0.41)	(0.39)	(0.38)	(0.46)	(0.46)
2 passengers	0.81	0.80	0.82	0.80	0.80
	(0.39)	(0.40)	(0.38)	(0.40)	(0.40)
3 passengers	0.19	0.20	0.18	0.20	0.20
	(0.39)	(0.40)	(0.38)	(0.40)	(0.40)
Females only	0.28	0.33	0.28	0.16	0.17
	(0.45)	(0.47)	(0.45)	(0.37)	(0.38)
Males only	0.29	0.26	0.29	0.36	0.39
	(0.46)	(0.44)	(0.46)	(0.48)	(0.49)
Females and males	0.43	0.41	0.43	0.48	0.44
	(0.50)	(0.49)	(0.50)	(0.50)	(0.50)
Young contestants (under 30)	0.38	0.39	0.39	0.36	0.32
	(0.49)	(0.49)	(0.49)	(0.48)	(0.47)
Middle age contestants (30 - 50)	0.41	0.41	0.41	0.44	0.49
	(0.49)	(0.49)	(0.49)	(0.50)	(0.51)
Old contestants (above 50)	0.04	0.03	0.05	0.04	0.05
	(0.19)	(0.16)	(0.22)	(0.20)	(0.22)
Contestants of different age	0.16	0.17	0.16	0.16	0.15
	(0.37)	(0.38)	(0.36)	(0.37)	(0.36)
Migration background	0.11	0.11	0.10	0.12	0.12
	(0.31)	(0.31)	(0.31)	(0.33)	(0.33)
DVD episode	0.09	0.08	0.05	0.18	0.22
	(0.29)	(0.28)	(0.22)	(0.39)	(0.42)
Observations	256	110	96	50	41

Source: Own calculations

Table 2 provides more detailed information, splitting the groups that reach the master question according to their initial opinion about playing it. These descriptives already point toward a certain degree of risk averse behavior by the contestants. While half of the groups are from the very beginning of the discussion against playing the master question (only one group then played nevertheless), only 5.5% directly agree to continue playing (all did so). Among the initially undecided groups, 63% finally decide to play the master question. In total, 49 groups (33.6% of those who reach the master question) take the risky option of either doubling their winnings or losing them completely.

With regard to group characteristics, the proportion of all-female groups is relatively low among the undecided groups. As far as age composition and migration background are concerned, no large differences emerge between the "Don't play" and "Undecided" groups. The composition of the "Play" group is somewhat different, but given the very low number of only 8 observations, these should not be over-interpreted. Not surprisingly, there are more undecided three-person groups than undecided two-person groups. It seems like the larger groups and those that are more diverse in terms of sex and age composition are less likely to have a uniform opinion.

Apart from the discussion time, the initial tendencies of the contestants and their respective share of the discussion (measured by the number of characters attributed to each contestant and the host from the discussion transcripts)[21] , the communication content was quantified by recording the main topics that were mentioned during the discussion. The six main topics were "modesty/humility", "nothing to lose", "a lot of money", encouragement or provocation by the host, and the estimated difficulty of the master question ("easy" or "difficult"). As soon as contestants talk about their modesty or humility in positive terms, this topic is coded as mentioned, i.e., the corresponding binary variable takes the value one. "Nothing to lose" is coded as mentioned if contestants say that even when answering incorrectly, they do not lose anything, which could be referred to as irrationality or a wrong reference point. "A lot of money" alludes to the contestants saying that the stake is already a significant sum of money for them. This is also coded if the contestants talk about important things they could buy with the stake. The fourth topic is only coded as mentioned if the host encourages the contestants to play the master question or provokes them. It is not coded when contestants themselves behave in this way. Finally, two variables indicate whether contestants talk about the anticipated difficulty of the master question. These are also coded if the contestants talk about their own high or low degree of knowledge. Table A in the appendix shows typical statements for each of the six topics.

Looking at the content of the discussion presented, all six topics are discussed by a significant share of groups, particularly among the undecided groups (see Table 2 and Table B in the appendix). The majority shares within the undecided groups indicate a tendency toward not playing the master question. Only 22% of these groups have a majority for playing, while 35% have an initial majority against playing. However, 63% of these groups actually continue to play (see Table 2).

As expected, the average discussion length is higher among the undecided groups (33.6 seconds) compared to the unanimous groups (22.2 seconds). Table 3 reports the results of descriptive OLS regressions of the (log of) discussion length on different sets of explanatory variables for all groups that have reached the master question (columns 1 - 6), and only the undecided ones (columns 7 - 12). For the whole sample, discussion length has a strong positive correlation with being undecided, and a slightly negative correlation with the winnings at stake in the most simple specification of column 1. These significant correlations disappear when indicators for initial majorities are included in the analysis.[22] Having a majority greatly shortens discussion time. For the sample of undecided groups, discussion time is shorter for the three-person groups which can be attributed to the possibility of a majority vote. This view is supported by the result that the significant correlation between discussion length and three-person groups vanishes with the inclusion of the majority indicators.

Looking at the discussion content, it can be seen that discussion length and the number of topics

[21] We also used the number of words instead of the number of characters to assess discussion shares. Both measures deliver the same results.

[22] This follows naturally since these indicator variables are highly collinear with the dummy variable for being undecided.

TABLE 2: Variable means (standard deviations)
by initial group opinion about master question

	All contestants	Initial opinion about master question		
		Don't play	Undecided	Play
Played master question	0.34	0.01	0.63	1.00
	(0.48)	(0.12)	(0.49)	(0.00)
Stake (/100)	8.12	8.38	7.92	7.38
	(2.46)	(2.57)	(2.24)	(3.06)
Stake per contestant (/100)	3.77	3.94	3.59	3.69
	(1.16)	(1.15)	(1.12)	(1.53)
Distance	4.72	4.78	4.64	4.82
	(1.32)	(1.42)	(1.21)	(1.25)
Number of questions	10.73	10.97	10.55	10.00
	(2.02)	(2.09)	(1.85)	(2.56)
Lives left	1.34	1.29	1.40	1.38
	(0.62)	(0.56)	(0.68)	(0.52)
Share of correct answers (in %)	84.38	84.26	84.67	83.07
	(6.24)	(5.58)	(7.02)	(5.76)
Share of known answers (in %)	65.38	64.78	66.31	63.36
	(11.25)	(10.34)	(11.85)	(14.75)
Streak of correct answers	2.72	2.74	2.78	2.00
	(2.70)	(2.36)	(3.16)	(1.07)
Wild card left	0.22	0.14	0.29	0.38
	(0.42)	(0.35)	(0.46)	(0.52)
2 passengers	0.82	0.86	0.74	1.00
	(0.39)	(0.35)	(0.44)	(0.00)
3 passengers	0.18	0.14	0.26	0.00
	(0.39)	(0.35)	(0.44)	(0.00)
Females only	0.24	0.29	0.15	0.50
	(0.43)	(0.46)	(0.36)	(0.53)
Males only	0.32	0.29	0.34	0.38
	(0.47)	(0.46)	(0.48)	(0.52)
Females and males	0.45	0.42	0.51	0.13
	(0.50)	(0.50)	(0.50)	(0.35)
Young contestants (under 30)	0.38	0.38	0.35	0.50
	(0.49)	(0.49)	(0.48)	(0.53)
Middle age contestants (30 - 50)	0.42	0.42	0.43	0.25
	(0.49)	(0.50)	(0.50)	(0.46)
Old contestants (above 50)	0.05	0.05	0.03	0.13
	(0.21)	(0.23)	(0.17)	(0.35)
Contestants of different age	0.16	0.14	0.18	0.13
	(0.37)	(0.35)	(0.39)	(0.35)
Migration background	0.11	0.11	0.12	0.00
	(0.31)	(0.31)	(0.33)	(0.00)
DVD episode	0.10	0.03	0.17	0.13
	(0.30)	(0.16)	(0.38)	(0.35)
Discussion time (in seconds)	27.25	22.14	33.58	22.50
	(21.06)	(15.00)	(25.55)	(15.30)
Humility/modesty	0.14	0.21	0.09	0.00
	(0.35)	(0.41)	(0.29)	(0.00)
A lot of money	0.16	0.19	0.15	0.00
	(0.37)	(0.40)	(0.36)	(0.00)
Nothing to lose	0.13	0.01	0.22	0.50
	(0.34)	(0.12)	(0.41)	(0.53)
Provocation/encouragement (by host)	0.33	0.32	0.38	0.00
	(0.47)	(0.47)	(0.49)	(0.00)
Easy to answer master question	0.08	0.03	0.12	0.25
	(0.28)	(0.16)	(0.33)	(0.46)
Difficult to answer master question	0.18	0.22	0.12	0.38
	(0.39)	(0.42)	(0.33)	(0.52)
Majority pro master	0.15	0.00	0.22	1.00
	(0.36)	(0.00)	(0.41)	(0.00)
Majority contra master	0.66	1.00	0.35	0.00
	(0.48)	(0.00)	(0.48)	(0.00)
Number of topics discussed	0.71	0.66	0.71	1.13
	(0.80)	(0.69)	(0.88)	(0.99)
Number of topics pro	0.21	0.04	0.34	0.75
	(0.46)	(0.20)	(0.54)	(0.71)
Number of topics contra	0.49	0.62	0.37	0.38
	(0.67)	(0.68)	(0.65)	(0.52)
Observations	146	73	65	8

Note: Majority shares do not sum up to 100% due to indifferent contestants.
Source: Own calculations

discussed are strongly positively correlated (even when controlling for initial majorities). One cannot infer from these results whether the discussion topics prolong the discussion, or whether a longer discussion gives rise to more topics. Likely, both mechanisms are present at the same time. Concerning the discussion topics, the activity of the host is significantly positively related to discussion length which suggests that he plays an active role and tries to influence the contestants. The mention of the valuation of the money at stake, and of the chances of a correct answer to the master question are also positively related to discussion length. Since these aspects are essential for the decision to be made, longer discussion can tentatively be seen as an indication of more "rational" decision making (see discussion below).

4 Empirical results

4.1 Econometric specification

"Quiz Taxi" allows one to study the decision of engaging in a risky activity, namely of choosing to play the master question instead of leaving the cab with the acquired winnings. From an expected utility point of view, contestants should determine their subjective probability P of correctly answering the master question and then continue to play if the expected utility from taking the gamble is higher than the utility they get from the winnings they have acquired so far:

$$P \cdot U(X_{play}, 2 \cdot stake) + (1 - P) \cdot U(X_{play}, 0) > U(X_{stop}, stake). \quad (1)$$

X_{play} and X_{stop} comprise factors other than the stake that are important for but unchanged by the decision, e.g., initial wealth, and factors that are changed by the decision, e.g., curiosity or reputation.[23] If contestants were risk neutral (and X_{play}, X_{stop} not too different), they would be indifferent between the two options if the probability of correctly answering the question was 50%. However, the data indicates that this probability is actually much higher: The average share of right answers is 84%, the average share of known answers is 65%, and 82% of those who play the master question answer correctly. Since only 34% of all groups play the master question (see Table 2), one can conclude that either contestants are risk averse or that they strongly underestimate their probability of being successful. The contestants of the US "Cash Cab" appear to be more risk loving: 45% of all groups go for the final "video bonus question" which is then answered correctly by 75% of the groups. The share of groups that make it to the final question is slightly higher in the US with 64% compared to 57%, too.

In order to empirically analyze the behavior of the "Quiz Taxi" contestants, the decision to play the master question is related to different sets of explanatory variables that refer to decision, contestant, and discussion characteristics. A similar approach is taken by Beetsma and Schotman (2001) and called

[23] For simplicity, no differentiation is made between factors that are changed by the decision and by success: e.g., there might be different reputation effects depending on not playing, playing and winning, and playing and losing.

TABLE 3: Estimation results for discussion length
(Coefficients from OLS regressions; all and undecided groups)

Dependent variable:	**All groups**						**Undecided groups only**					
Log of discussion time	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
Undecided to play master question	0.269**	0.071	0.314***	0.115	0.299***	0.103	–	–	–	–	–	–
	(0.119)	(0.136)	(0.112)	(0.124)	(0.088)	(0.114)						
Stake (/100)	-0.053**	-0.048**	-0.033	-0.029	-0.007	-0.003	-0.026	-0.009	0.004	0.025	0.008	0.031
	(0.024)	(0.023)	(0.023)	(0.023)	(0.019)	(0.018)	(0.043)	(0.039)	(0.038)	(0.033)	(0.035)	(0.029)
Share of correct answers (in %)	0.016	0.012	0.009	0.005	-0.003	-0.008	0.004	-0.003	0.002	-0.008	-0.011	-0.020*
	(0.011)	(0.011)	(0.010)	(0.010)	(0.008)	(0.008)	(0.016)	(0.015)	(0.013)	(0.012)	(0.011)	(0.011)
Wild card left	0.007	0.063	0.061	0.109	0.214*	0.246**	0.160	0.242	0.232	0.352**	0.344**	0.438***
	(0.139)	(0.130)	(0.130)	(0.125)	(0.117)	(0.117)	(0.192)	(0.179)	(0.173)	(0.163)	(0.154)	(0.140)
3 passengers	-0.095	-0.010	-0.132	-0.048	-0.169	-0.107	-0.363**	-0.239	-0.365**	-0.239	-0.389**	-0.299*
	(0.143)	(0.145)	(0.126)	(0.126)	(0.105)	(0.104)	(0.169)	(0.173)	(0.158)	(0.154)	(0.154)	(0.153)
Females only	-0.132	-0.086	-0.104	-0.059	-0.045	-0.022	-0.054	0.016	-0.002	0.109	0.003	0.078
	(0.139)	(0.136)	(0.127)	(0.123)	(0.110)	(0.106)	(0.239)	(0.225)	(0.211)	(0.188)	(0.175)	(0.146)
Males only	0.060	0.136	0.045	0.121	-0.034	0.013	-0.080	0.068	-0.059	0.113	-0.136	-0.013
	(0.121)	(0.116)	(0.111)	(0.107)	(0.094)	(0.089)	(0.174)	(0.164)	(0.158)	(0.150)	(0.145)	(0.142)
DVD episode	0.480**	0.314*	0.251	0.112	0.057	-0.053	0.702***	0.476**	0.279	0.029	0.165	-0.049
	(0.195)	(0.175)	(0.178)	(0.156)	(0.144)	(0.138)	(0.202)	(0.196)	(0.203)	(0.177)	(0.164)	(0.157)
Majority pro master	–	-0.649***	–	-0.593***	–	-0.324**	–	-0.617***	–	-0.519**	–	-0.252
		(0.192)		(0.172)		(0.160)		(0.212)		(0.208)		(0.193)
Majority contra master	–	-0.519***	–	-0.546***	–	-0.503***	–	-0.516***	–	-0.673***	–	-0.605***
		(0.175)		(0.187)		(0.184)		(0.183)		(0.216)		(0.220)
Number of topics pro	–	–	0.316***	0.232**	–	–	–	–	0.390***	0.239	–	–
			(0.096)	(0.112)					(0.126)	(0.145)		
Number of topics contra	–	–	0.326***	0.344***	–	–	–	–	0.315**	0.442***	–	–
			(0.087)	(0.083)					(0.138)	(0.135)		
Humility/modesty	–	–	–	–	0.055	0.123	–	–	–	–	-0.316	-0.020
					(0.149)	(0.147)					(0.200)	(0.216)
A lot of money	–	–	–	–	0.302**	0.380***	–	–	–	–	0.322	0.506**
					(0.126)	(0.125)					(0.223)	(0.199)
Nothing to lose	–	–	–	–	0.240**	0.081	–	–	–	–	0.196	0.018
					(0.112)	(0.130)					(0.161)	(0.169)
Provocation/encouragement (by host)	–	–	–	–	0.683***	0.666***	–	–	–	–	0.652***	0.569***
					(0.084)	(0.087)					(0.127)	(0.118)
Easy to answer master question	–	–	–	–	0.257*	0.237	–	–	–	–	0.332*	0.362*
					(0.151)	(0.155)					(0.188)	(0.201)
Difficult to answer master question	–	–	–	–	0.552***	0.531***	–	–	–	–	0.612**	0.557*
					(0.124)	(0.129)					(0.242)	(0.286)
Constant	2.053**	2.753***	2.203***	2.990***	2.831***	3.663***	3.117**	3.790***	2.815***	3.745***	3.750***	4.542***
	(0.879)	(0.886)	(0.786)	(0.847)	(0.644)	(0.752)	(1.219)	(1.188)	(0.992)	(1.039)	(0.928)	(0.965)
Adjusted R2	0.12	0.18	0.24	0.30	0.45	0.48	0.12	0.24	0.25	0.39	0.44	0.52
Model p-value	0.001	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000
Observations	146	146	146	146	146	146	65	65	65	65	65	65

Notes: Table reports coefficients of OLS estimations of the log of discussion time (in seconds) on specified variables, robust standard errors in parentheses. Significance levels: *10% **5% ***1%
Source: Own calculations

an "explanatory data analysis" with "rules-of-thumb variables" (Beetsma and Schotman 2001, section 3). The general regression framework can be written as:

$$P(y_i = 1|X_i) = G(X_i, \beta), \tag{2}$$

where y_i is an indicator variable that takes on the value 1 if group i decides to play the master question, and the value 0 otherwise. The vector X_i comprises different sets of the variables presented in Table 2. These variables include, e.g., the winnings at stake, the share of correct or known answers, sex and age composition of the group, as well as discussion length and topics.

The function G relates the explanatory variables to the final decision. The most obvious choice for the estimation model is a standard probit model $G(X_i, \beta) = \Phi(X_i'\beta)$ where Φ denotes the cumulative distribution function of the standard normal. As robustness checks, Table C in the appendix shows results from a linear probability model (OLS-estimation) and a logit model for a baseline specification. The results are robust to the choice of the estimation model.[24]

The main sample that is used for the following analysis consists of the 146 groups which reach their destination and are thus asked if they want to play the master question. A group of special interest consists of the 65 groups that are initially undecided whether to play the master question or not.

As discussed in Section 3, the "Quiz Taxi" contestants might represent a selected sample of rather risk-loving individuals. While it is impossible to control for this selection process, it is possible to explicitly control for the selection process that takes place during the cab ride. As argued by Bliss et al. (2011), it is unlikely that this selection is directly based on risk attitudes. But since only contestants who did not answer more than two questions incorrectly reach the master question, they might over-represent people with a higher level of knowledge who would find it easier to answer the master question. Hence, one could presume an upward bias on the coefficient of the money at stake. The data allows one to control for this possibility by estimating a Heckman selection model. The instrument used in the first stage regression for reaching the master question is the distance from the start of the cab ride to the final destination. This information can be confidently regarded as unrelated to the knowledge or any other characteristics of the contestants because the potential candidates state their destination before the nature of the show is revealed to them.[25] In contrast, the distance to the destination should significantly reduce the chances of reaching it because the longer the distance, the more questions posed, and the higher the probability of getting three questions wrong. The results of column 2 in Table C in the appendix strongly support this view: A 1 kilometer greater distance bears a highly significant relation to an approximately 8.5%-points lower probability of reaching the master question. The results of Table C also show that the bias arising

[24] Since results diverge a little bit for more extensive variable specifications and particular samples between the linear (OLS) and the non-linear models (probit and logit), marginal effects from the "correct" and more intuitive probit model are reported in the subsequent Tables. No difference between results from probit or logit models arise in any specification.

[25] According to two contestants from the show, the destination was actually determined by the producers. In this case, it is definitely exogenous to the contestants.

from this selection process can be adequately controlled for by adding the share of correct answers as a confounder to proxy for the knowledge and confidence of the contestants.

4.2 Baseline analysis

The first analysis only rests on readily observable characteristics of the decision and the group: the stake, the share of right answers given by the group during the ride, a dummy variable indicating whether the group still has a wild card left (which can then be used for the master question), the group's size and sex composition, as well as an indicator variable for rides from the DVD's. With regard to risk attitude, the stake is the most important variable. If contestants were risk averse, the probability of playing the master question would decline as the stake rises.

Equation 1 illustrates that the group's decision to play the master question does not only depend on risk attitude, but also critically on the subjective probability of correctly answering the master question which is unobservable. In the estimations, the share of correct answers to the previous questions is used as a proxy variable. This variable also controls for the possibility that smarter contestants have acquired more money, which could bias the coefficient of the stake upward if omitted (see discussion above). This variable probably also captures other things than knowledge, e.g., confidence that has been established during the cab ride. It is chosen over other possibilities (share of known answers, or measures only based on questions worth more than EUR 100) because it is easy to remember for the contestants and the most objective measure.[26] In addition, the "Wild card left"-dummy also proxies for the group's knowledge because smarter groups are less likely to have used wild cards. A wild card should also increase the probability to play the master question because more persons are allowed to answer.

The results of the baseline estimations are reported in the first 2 columns of Table 4 and Table 5 for all groups and the undecided groups, respectively. They clearly indicate risk averse behavior by the contestants. While the raw correlation between the stake and playing the master question is negative but statistically insignificant (column (1)), the correlation becomes significantly negative with the inclusion of control variables (column (2) of Table 4). In the full sample, an increase in the winnings at stake by EUR 100 is associated with a decline in the probability of playing the master question by about 4%-points, which equals more than 10% of the average sample probability of playing (34%). As expected, the share of correct answers bears a significantly positive relation to the probability of playing the master question. On average, 10 questions are played during the cab ride (see Table 2). Hence, one more correct answer translates into a 13%-points higher probability of continuing to play. Comparing the results for all and undecided groups, it is found that the stake matters more in the full sample and the share of correct answers more for undecided groups. This suggests that the formation of initial opinions is dominated by the money at stake, but that probabilities become more important in discussions.

[26] In assessing which question was known or guessed, some subjectivity enters the data. All of these measures are also strongly positively correlated with each other.

Most of the other included variables are not significantly related to the group's decision.[27] This is somewhat surprising for the indicator variables for having a wild card left, and being a group of three. Both can be expected to have a positive influence because they increase the probability of being able to answer the master question by enlarging the group of people who are allowed to answer. In addition, being three rather than two persons reduces the money at stake per contestant and thus personal risk (assuming that winnings are shared equally). In contrast, Bliss et al. (2011) report that three and four person groups choose to play the final question more often than individuals or two person groups. They also argue that the group decision is based on the overall amount of money at stake, and not on the stake per contestant.

The dummy variable for DVD episodes is significant in the full sample which provides evidence that these rides are indeed strongly selected, but only half in size and insignificant for undecided groups. Among the undecided, all-female groups choose significantly less often to play the master question which is in-line with the finding of Bliss et al. (2011) and Jianakoplos and Bernasek (1998) that women are more risk averse than men. The findings of Dohmen et al. (2009a), Fehr-Duda et al. (2006) and Fehr-Duda et al. (2011) suggest that this result is not only due to larger risk aversion, but also due to women weighting probabilities differently than men.

Tables D and E in the appendix provide several robustness checks concerning the variables included in the regressions. Adding the age composition and the cultural background of the group does not affect any estimates.[28] Using the stake per contestant, the log of the stake, or adding an interaction of the stake and the share of correct answers does not substantially change any findings discussed above. However, it appears that a positive effect of a three person group is rather due to less risk per contestant than to an increased probability of answering correctly. Concerning the influence of the group size on playing the master question, one also has to bear in mind the internal decision process in the group. Assuming that a consensus decision must be reached, otherwise the master question will not be played, a larger group leads to a lower probability of playing the master question. This is due to the presence of an additional "veto player" who can prevent the choice of the risky alternative.

Using the share of known answers[29] or the final streak of consecutive correct answers before the master question as proxies for the knowledge of the group does not substantially alter any of the findings, either. The respective correlations remain positive but less strong, and are only statistically significant for the undecided groups. This suggests that it is likely not only knowledge, but also confidence and enthusiasm that positively influences the decision to take the risky alternative. In that sense, the final decision might

[27] The large number of insignificant coefficients explains the high p-value of the models. However, the main interest here is not on the model as a whole but on the correlations between single variables and the group's decision, i.e., on the corresponding t-test which have a much higher power than the F-test for the overall model significance.

[28] These are the variables that potentially suffer most from measurement error because age and migration background must be guessed in many cases. Given the small sample size, they are excluded from all regressions in order to keep the model as parsimonious as possible.

[29] Known answers are all correct answers that were given without the help of a wild card, or that were obvious guesses. Some subjectivity enters here because one has to decide if the answer was a guess or not. In addition, wild cards are also used in cases where the answer is actually known but contestants are hesitant and want to back up their decision.

be somewhat "path-dependent" (see Post et al. 2008, for a discussion of path-dependency in the show "Deal or No Deal").

For illustrative purposes, the relative risk aversion parameter γ of the standard CRRA utility function $U(stake) = \frac{stake^{1-\gamma}}{1-\gamma}$ is estimated via maximum likelihood estimation for the probability of playing the master question.[30] This estimation is complicated by censoring issues - contestants can only bet all of their acquired winnings or nothing (see Gertner 1993, Metrick 1995, Bliss et al. 2011) and the unknown subjective probabilities for answering the master question correctly. Using the share of *correct* answers as a proxy for this probability, γ is estimated at 4.8.[31] Using the share of *known* answers, γ is estimated at 2.4. For a uniform probability of 50% for all contestants, γ is estimated at 1.2. All of these estimates are based on the assumption of very low wealth (EUR 20 which can be seen as the value of the cab ride) and thus present conservative estimates of risk aversion.[32] For the more realistic assumption of larger wealth, the risk aversion coefficients would be even higher (e.g., 4 to 14 for EUR 1,000). In a very similar setting, Beetsma and Schotman (2001) estimate a risk aversion coefficient of 0.42 (Beetsma and Schotman 2001, Table). Taking this value and estimating the corresponding expected probability for a correct answer gives an estimate of 24%. By and large, our estimates are in-line with the evidence presented by Beetsma and Schotman (2001). In any case, the hypothesis of risk neutrality is clearly rejected, even if people would believe that their chances of answering the master question are very low. These results also accord reasonably well with those of Dohmen et al. (2009b) based on large experimentally-validated microeconomic survey data. Compared to other game show studies like Hartley et al. (2006) or Hersch and McDougall (1997), the estimates presented here are somewhat higher, which could be due to smaller selection effects at work in "Quiz Taxi", as discussed in section 3.

For comparison with the results of Bliss et al. (2011) for the US "Cash Cab", coefficients of absolute risk aversion α are also estimated ($U(stake) = 1 - \exp(-\alpha \cdot stake)$). Depending on the proxy for the subjective probabilities, α hovers between 0.003 and 0.009. In line with the descriptives discussed above, the contestants of the German version of the show display a much larger degree of risk aversion. These estimates are also considerably higher than those reported by, e.g, Gertner (1993) or Metrick (1995), but fits in with the coefficients presented in Table 5 in Cohen and Einav (2007).[33]

[30] The CRRA point estimates should not be over-interpreted since they are based on a sample of only 146 independent groups. We thank our colleague Maarten van Kampen for help with the maximum likelihood procedure, and Peter Schotman and Philip Hersch for providing us with the codes that they used in their papers.

[31] Using (out-of-sample) predictions for the probability to answer the master question correctly obtained from probit regressions with the sample of groups that play the master questions gives a similar, but very imprecise estimate.

[32] Another argument for setting the reference wealth level to zero or equal to the acquired winnings comes from prospect theory (Kahneman and Tversky 1979) where individuals are assumed to evaluate gains and losses directly and not with regard to their overall wealth (see also Beetsma and Schotman 2001, Fullenkamp et al. 2003).

[33] Clearly, the comparability of all of these estimates is very limited because they are obtained from very different samples using various methodologies. Therefore, the external validity of estimates of risk aversion parameters is highly questionable (see Rabin 2000).

TABLE 4: Estimation results for decision to play the master question
(Average marginal effects from probit regressions; all groups)

Dependent variable: **Play master question (1=yes, 0=no)**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Stake (/100)	-0.024	-0.044***	-0.036**	-0.020	-0.025**
	(0.016)	(0.016)	(0.016)	(0.012)	(0.011)
Share of correct answers (in %)	–	0.013**	0.010	0.011**	0.013***
		(0.006)	(0.007)	(0.005)	(0.004)
Wild card left	–	0.075	0.064	0.057	-0.018
		(0.093)	(0.093)	(0.079)	(0.081)
3 passengers	–	0.071	0.091	0.114	0.148**
		(0.106)	(0.102)	(0.074)	(0.064)
Females only	–	-0.138	-0.118	-0.132	-0.142**
		(0.095)	(0.091)	(0.081)	(0.066)
Males only	–	-0.033	-0.023	-0.025	-0.015
		(0.090)	(0.087)	(0.062)	(0.060)
DVD episode	–	0.270**	0.173	0.083	0.146
		(0.129)	(0.135)	(0.110)	(0.110)
Discussion time (in seconds)	–	–	0.005***	0.006***	0.009***
			(0.002)	(0.002)	(0.002)
Number of topics pro	–	–	–	0.306***	–
				(0.061)	
Number of topics contra	–	–	–	-0.297***	–
				(0.056)	
Humility/modesty	–	–	–	–	-0.696***
					(0.109)
A lot of money	–	–	–	–	-0.476***
					(0.095)
Nothing to lose	–	–	–	–	0.465***
					(0.064)
Provocation/encouragement (by host)	–	–	–	–	-0.225***
					(0.058)
Easy to answer master question	–	–	–	–	0.058
					(0.097)
Difficult to answer master question	–	–	–	–	-0.244***
					(0.077)
Pseudo R2	0.01	0.09	0.13	0.41	0.54
Model p-value	0.129	0.027	0.008	0.000	0.000
Observations	146	146	146	146	146

Notes: Table reports average marginal effects from probit regressions, robust standard errors in parentheses. In specification (5), 4 failures and 1 success are completely determined. Significance levels: *10% **5% ***1%
Source: Own calculations

TABLE 5: Estimation results for decision to play the master question
(Average marginal effects from probit regressions; undecided groups only)

Dependent variable:					
Play master question (1=yes, 0=no)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Stake (/100)	-0.005	-0.044	-0.041	-0.031*	-0.054***
	(0.028)	(0.030)	(0.029)	(0.017)	(0.016)
Share of correct answers (in %)	–	0.030**	0.028**	0.015**	0.024***
		(0.013)	(0.012)	(0.007)	(0.007)
Wild card left	–	-0.083	-0.091	-0.064	-0.152*
		(0.131)	(0.124)	(0.096)	(0.087)
3 passengers	–	0.019	0.093	0.192**	0.269***
		(0.128)	(0.123)	(0.093)	(0.102)
Females only	–	-0.299**	-0.282**	-0.344***	-0.318***
		(0.136)	(0.142)	(0.098)	(0.091)
Males only	–	-0.112	-0.089	-0.140*	-0.131
		(0.130)	(0.123)	(0.084)	(0.081)
DVD episode	–	0.132	0.005	0.137	0.231**
		(0.164)	(0.176)	(0.148)	(0.099)
Discussion time (in seconds)	–	–	0.006**	0.007***	0.012***
			(0.003)	(0.002)	(0.002)
Number of topics pro	–	–	–	0.228**	–
				(0.113)	
Number of topics contra	–	–	–	-0.373***	–
				(0.071)	
Humility/modesty	–	–	–	–	-0.515***
					(0.098)
A lot of money	–	–	–	–	-0.576***
					(0.113)
Provocation/encouragement (by host)	–	–	–	–	-0.271***
					(0.088)
Easy to answer master question	–	–	–	–	-0.087
					(0.134)
Difficult to answer master question	–	–	–	–	-0.404***
					(0.152)
Pseudo R2	0.00	0.14	0.19	0.50	0.56
Model p-value	0.870	0.209	0.112	0.000	0.000
Observations	65	65	65	65	65

Notes: Table reports average marginal effects from probit regressions, robust standard errors in parentheses. In specification (5), 1 failure is completely determined. "Nothing to lose" not included because of perfect prediction of success. Significance levels: *10% **5% ***1%
Source: Own calculations

4.3 Communication analysis

While it is certainly interesting to deduce risk attitudes by looking at the relationship between choosing to play the master question and the winnings at stake or the prior performance on the show, the "Quiz Taxi" offers much more information about the decision making process of the group. As discussed in section 3, one can observe the communication within the group, and thus gain information about the length of the discussion, the arguments exchanged, and initial opinions about the choice to be made. By doing so, one can to some extent open the "black box" of the decision making-process: Variables that typically remain hidden to researchers can be included in the analysis. The indicator variables for the discussion topics can be seen as proxies for some of the factors X_{stop} and X_{play} in Equation 1, e.g., wealth effects or personality traits. Omitting such variables could bias estimation results.

Columns (3) to (5) of Tables 4 and 5 report the results from regressions that incorporate the additional information obtained from the group discussions for all and undecided groups. The latter especially have a strong need for communication in order to come to a mutual agreement. On average, their discussions last about 11 seconds, i.e., 50%, longer (see Table 2). For both samples, it is not necessarily the case that the arguments are exchanged first and then the decision is made. It is possible that the decision has already been made and the arguments mentioned only serve to support this decision. This does not pose a problem because the interest is on the observation that the decision to play the master question

is associated with certain traits, beliefs or motives of the contestants. It must only be assumed that the contestants do not intentionally lie about their motives which seems very unlikely. For an overview of which groups discuss which topics, see Table 2; Table B in the appendix gives additional information about the discussion topics.

The inclusion of the discussion topics and the discussion length greatly improves the overall model fit. Of all 146 groups (65 undecided groups), 102 (44) discuss at least 1 of the topics. Most of the discussion topics bear a strongly significant relation to the decision made and all of them point into the direction that one would expect. In this sense, decisions made by the "Quiz Taxi" contestants can be seen as subjectively rational because they do what they believe to be right.

Groups that say that they are humble are much less likely to play, the same is true for groups that regard their acquired winnings as a lot of money. Those who believe that they cannot lose anything are very likely to play the master question. All of these correlations are much stronger for the sample of undecided groups than for the whole sample.[34]

The discussion length always has a strongly positive correlation with the decision to play the master question, which might be due to contestants realizing after some time that playing the master question is an attractive option or that a longer discussion may introduce another factor that also influences the decision.[35] Table 3 shows that the discussion length is strongly related to the activity of the host, and the topics discussed. The activity of the host is likely endogenous, i.e., he provokes or encourages contestants that do not want to play. Omitting discussion time from the regressions, no significant correlation between the activity of the host and the final decision is found. Therefore, it can be speculated that the host makes "unwilling" contestants discuss their decision longer, and that the longer discussion increases the likelihood of choosing to play the master question, which could be seen as the "rational" decision as discussed in section 3.2. In this case, the host would indirectly be able to talk the contestants into taking the risky alternative.

Regarding the main variables of interest from the baseline regressions, some differences emerge once the discussion characteristics are included. The stake is still negatively correlated with playing the master questions, but significance is reduced, except for the most extensive specifications in columns (5). It now also appears to be more important among the undecided groups. The share of correct answers is now significantly positively associated with playing the master question in all but one specification. This relation remains much stronger for the undecided groups.

In contrast to the previous regressions, it now also seems as if all-female groups are much less likely to play, especially when they are initially undecided. In the full sample, all-female groups have a 14%-points lower propensity to play the master questions, which even decreases to -32%-points for the undecided

[34] Having nothing to lose perfectly predicts that an initially undecided group will play, and is thus not included in the regressions. These coefficients are likely inflated due to the small sample size.

[35] This significant correlation also holds if the longest discussions are excluded from the analysis, or if the log of discussion time is used (see Tables D and E).

groups. In the most extensive specifications, three person groups are now much more likely to play than two person groups by approximately 15%-points, and by 27%-points among the undecided.

The very high R-squared of the discussion characteristics-augmented regressions might be regarded as problematic. Therefore, all regressions are also carried out including each discussion topic separately (the dummies for the difficulty of the master question are included together).[36] The results of these regressions are reported in Table 6. This also allows one to see which variables impact individually on the other explanatory variables, i.e., omitting which variables would lead to an estimation bias. Most importantly, none of the previously discussed findings change substantially.

It stands out that the discussion topics that refer to the valuation of the money at stake impact most on the other coefficients. Including the "Nothing to lose"-dummy in the regressions adds most explanatory power.[37] "Nothing to lose" is negatively correlated with the stake: Groups who believe that there is nothing to lose have acquired EUR 100 less on average (see Table B). Hence, the change in the coefficient of the stake can be seen as further evidence for increasing risk aversion with regard to the winnings at stake. But it could also suggest that failing to control for reference points results in a downward bias on the estimated coefficient of the money at stake. In contrast, the "A lot of money"-dummy, which could be expected to have similar effects because it can be seen as the opposite end of the same aspect, impacts on the estimation results rather in the opposite direction, i.e., omitting this variable creates an upward bias on "Stake".

Except for the "DVD's"-dummy, none of the other coefficients change considerably with the inclusion of the discussion topics. The changes in the coefficient of the "DVD's"-dummy further support the view that these rides represent a very selected sample of particularly interesting contestants.

So far, only the determinants of the decision to play the master question have been studied. The question whether the final decision is a good one remains open, and would require knowledge of the unknown counter-factual situation. That is, would groups that decline to play answer the master question correctly? Would groups that play the master question be better off if they had stopped? The latter question can be partly addressed by looking at the performance of the 50 groups that play the master question. The small number of observation limits the possibility of a multivariate analysis, especially because only 9 groups answer incorrectly. Table 7 compares variable means for the groups that answer correctly or wrongly to the master question.

82% of these groups answer correctly. Therefore, playing the master question can be seen as an attractive option. Those who give the right answer, have on average acquired more money. The difference is statistically significant between the two groups at a 10% significance level and suggests that groups with more money at stake make the better decision. If risk averse, a higher stake should lead contestants to think more carefully and to continue only if the chance of answering correctly are high. A similar

[36] The mutual correlations of the discussion topics are quite low.

[37] In the sample of undecided groups, "Nothing to lose" perfectly predicts that a group plays the master question.

TABLE 6: Estimation results: Each topic separately
(Average marginal effects from probit regressions; all and undecided groups)

	All groups					Undecided groups only			
Dependent variable:	**Humility/**	**A lot of**	**Nothing**	**Provo-**	**Difficulty**	**Humility/**	**A lot of**	**Provo-**	**Difficulty**
Play master question (1=yes, 0=no)	**modesty**	**money**	**to lose**	**cation**	**of master**	**modesty**	**money**	**cation**	**of master**
Stake (/100)	-0.035**	-0.043***	-0.023	-0.040***	-0.032**	-0.048*	-0.053**	-0.050*	-0.043
	(0.015)	(0.015)	(0.014)	(0.015)	(0.016)	(0.026)	(0.025)	(0.026)	(0.029)
Share of correct answers (in %)	0.008	0.015**	0.009*	0.012*	0.010	0.023**	0.028***	0.033***	0.027**
	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.010)	(0.010)	(0.010)	(0.011)
Wild card left	0.069	0.015	0.061	0.020	0.064	-0.104	-0.081	-0.138	-0.118
	(0.091)	(0.090)	(0.084)	(0.092)	(0.092)	(0.108)	(0.098)	(0.120)	(0.122)
3 passengers	0.101	0.133	0.096	0.124	0.085	0.121	0.233**	0.145	0.112
	(0.090)	(0.090)	(0.090)	(0.099)	(0.099)	(0.108)	(0.098)	(0.115)	(0.128)
Females only	-0.140	-0.110	-0.131*	-0.136	-0.114	-0.291**	-0.253**	-0.291**	-0.294**
	(0.087)	(0.087)	(0.075)	(0.088)	(0.090)	(0.121)	(0.126)	(0.136)	(0.136)
Males only	-0.029	-0.000	-0.066	0.006	-0.021	-0.080	-0.164*	-0.037	-0.074
	(0.081)	(0.079)	(0.078)	(0.087)	(0.086)	(0.106)	(0.096)	(0.118)	(0.120)
DVD episode	0.228*	0.246**	0.120	0.200	0.127	0.185	0.228*	0.023	0.009
	(0.133)	(0.121)	(0.124)	(0.130)	(0.136)	(0.197)	(0.127)	(0.161)	(0.180)
Discussion time (in seconds)	0.005***	0.006***	0.004**	0.008***	0.006***	0.007***	0.008***	0.010**	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)
Humility/modesty	-0.670***	–	–	–	–	-0.803***	–	–	–
	(0.140)					(0.194)			
A lot of money	–	-0.515***	–	–	–	–	-0.693***	–	–
		(0.115)					(0.152)		
Nothing to lose	–	–	0.572***	–	–	–	–	–	–
			(0.112)						
Provocation/encouragement (by host)	–	–	–	-0.229**	–	–	–	-0.253**	–
				(0.090)				(0.127)	
Easy to answer master question	–	–	–	–	0.157	–	–	–	-0.063
					(0.136)				(0.193)
Difficult to answer master question	–	–	–	–	-0.177*	–	–	–	-0.255
					(0.103)				(0.182)
Pseudo R2	0.23	0.22	0.29	0.16	0.15	0.33	0.36	0.24	0.22
Model p-value	0.000	0.004	0.000	0.003	0.003	0.017	0.053	0.131	0.091
Observations	146	146	146	146	146	65	65	65	65

Notes: Table reports average marginal effects from probit regressions, robust standard errors in parentheses. "Nothing to lose" perfectly predicts success for undecided groups. Significance levels: *10% **5% ***1%
Source: Own calculations

TABLE 7: Variable means by answer to master question

	Correct answer	Wrong answer
Stake (/100)	7.87	7.00
Stake per contestant (/100)	3.66	3.20
Share of correct answers (in %)	85.61	84.14
Share of known answers (in %)	66.70	66.51
Streak of correct answers	3.20	2.33
Wild card left	0.29	0.33
3 passengers	0.20	0.22
Females only	0.17	0.11
Males only	0.39	0.22
Discussion time (in seconds)	38.12	28.00
Humility/modesty	0.02	0.00
A lot of money	0.07	0.00
Nothing to lose	0.34	0.44
Provocation/encouragement (by host)	0.37	0.22
Easy to answer master question	0.20	0.00
Difficult to answer master question	0.17	0.11
Majority pro master	0.41	0.44
Majority contra master	0.10	0.00
Number of topics discussed	0.80	0.56
Number of topics pro	0.54	0.44
Number of topics contra	0.27	0.11
Observations	41	9

Source: Own calculations

argument is made by Fehr-Duda et al. (2010) who attribute increasing relative risk aversion when stakes rise to individuals weighing probabilities more rationally. In contrast, none of the other readily observable characteristics do not differ by much between the two groups.

Looking at the discussion-related characteristics, some more differences emerge. On average, groups that answer correctly have discussed their decision more than 10 seconds longer (difference at the edge of statistical significance at 10% level), and have exchanged more arguments (not statistically significantly different). Looking at the discussion topics also supports the view that carefully thinking about the decision pays off. None of the groups that mention "Humilty/modesty" (1 group only), "A lot of money" (3 groups) or "Easy to answer" (8 groups) and only 1 of 8 groups that mention "Difficult to answer" answer incorrectly. And only 2 of 15 groups that were provoked or encouraged by the host get the master question wrong. However, 4 of 18 groups that believe that there is nothing to lose, which is not a really good argument for playing, give the wrong answer. In sum, the evidence suggests that the discussions indeed greatly help the contestants to make the right decision.

5 Conclusion

This paper contributes to the literature on risk attitudes of groups by using data from the German TV game show "Quiz Taxi". After a cab ride during which contestants can earn money by answering knowledge questions, they are asked if they want to play a final master question with which they can double their winnings, or lose everything.

"Quiz Taxi" has two main features that make it very attractive for studying risk attitudes. Firstly, the group of contestants represents a much less selected sample than in other TV game shows because the contestants do not actively apply to be on the show and do not play in front of a large TV studio

audience. Secondly, the "Quiz Taxi" contestants always play in groups of two or three people. This makes it possible to observe the communication between group members, and to open the "black box" of decision making to some extent.

Overall, contestants show fairly risk averse behavior. The risk parameter of a standard CRRA-utility function is estimated at rather high values of approximately 1 to 5 for reasonable subjective probabilities of being able to answer the master question. Compared to other studies using game show data, the analysis supports findings of, e.g., Beetsma and Schotman (2001) or Dohmen et al. (2009b) of rather high but not unreasonable degrees of risk aversion which is likely due to smaller selection effects in the setup presented here. Still, there is reason to believe that the Quiz Taxi contestants are on average more risk-loving than the average population. Hence, the evidence presented should be regarded as conservative estimates of risk aversion.

The regression analysis suggests that an increase in the winnings at stake by EUR 100 (the average value of one question) is associated with a decline in the probability of playing the master question by approximately 4%-points on average. A higher share of correct answers during the preceding cab ride of 10%-points (1 more correct answer) is associated with an increase in the probability of playing by about 13%-points. This positive relationship can be attributed to greater knowledge and/or more confidence of the successful contestants. It is also shown that all-female groups are less likely to choose the risky alternative, and that three-person groups are slightly more risk-loving than two-person groups. Both findings are also reported by Bliss et al. (2011) who analyze risk attitudes with data from the US version of the show called "Cash Cab". While it remains unclear whether women are really more risk averse or underestimate their chances of answering correctly, the analysis suggests that for three person groups the lower risk per contestant weighs more heavily than the higher probability of a right answer for the final decision.

The analysis of the communication characteristics and its content shows that contestants typically make consistent decisions in the sense that they have arguments to back up their decision, but not necessarily in the sense that the set of arguments discussed is exhaustive. Not playing the master question is related to arguments of humility or modesty, the difficulty of the final question, or the large amount of money at stake. Playing the master question is related to arguments of having nothing to lose (a different reference point), or the easiness of the final question. It is shown that the inclusion of these characteristics matters for the estimation of the other coefficients. In particular, failing to account for the individual valuations of the money at stake creates an omitted variables pertaining to the coefficients of the money at stake.

The analysis also reveals that it is important to differentiate between groups that are immediately determined to stop or to continue playing (90% of these groups stop), and groups that are initially undecided. The decision to play is positively related to the discussion length, especially for initially

undecided groups, which might induce more rationality into the decision making process. While the initial tendency is dominated by risk aversion with regard to the money at stake, a longer discussion brings about more arguments and enables contestants to make better informed decisions.

The behavioral analysis highlights the importance of people's beliefs, expectations, and values for decision making processes in situations involving risky choices. Future research should account for such typically unobserved personal factors, especially in the context of group decision making where arguments must be exchanged in order to reach a mutual agreement. Since the "Quiz Taxi" has been aired in many countries since 2005, this show also represents an opportunity to study cultural differences in decision making behavior under uncertainty. Compared to their US counterparts (Bliss et al. 2011), the German contestants seem to be more risk averse at first sight.

References

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutstroem, "Risk Aversion in Game Shows," in Glenn W. Harrison James C. Cox, ed., *Risk Aversion in Experiments. Research in Experimental Economics*, Bingley, UK: Emerald Publ., 2008, pp. 359–404. Vol. 12.

Attanasio, Orazio P. and Guglielmo Weber, "Intertemporal Substitution, Risk Aversion and the Euler Equation for Consumption," *The Economic Journal*, 1989, *99* (395), 59–73.

Baker II, Ronald J., Susan K. Laury, and Arlington W. Williams, "Comparing Small-Group and Individual Behavior in Lottery-Choice Experiments," *Southern Economic Journal*, October 2008, *75* (2), 367–382.

Barsky, Robert B., "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," *The Quarterly Journal of Economics*, 1997, *112* (2), 537–79.

Beetsma, Roel M. W. J. and Peter C. Schotman, "Measuring Risk Attitudes in a Natural Experiment: Data from the Television Game Show Lingo," *The Economic Journal*, October 2001, *111* (474), 821–848.

Blinder, Alan S., "Monetary policy by committee: Why and how?," *European Journal of Political Economy*, March 2007, *23* (1), 106–123.

Bliss, Richard T., Mark E. Potter, and Christopher Schwarz, "Decision Making and Risk Aversion in the Cash Cab," Working Paper Series, SSRN eLibrary 2011.

Brooks, Robert, Robert Faff, Daniel Mulino, and Richard Scheelings, "Deal or No Deal, That is the Question: The Impact of Increasing Stakes and Framing Effects on Decision-Making under Risk," *International Review of Finance*, 2009, *9* (1-2), 27–50.

Cohen, Alma and Liran Einav, "Estimating Risk Preferences from Deductible Choice," *American Economic Review*, June 2007, *97* (3), 745–788.

Cooper, David J. and John H. Kagel, "Are Two Heads Better than One? Team versus Individual Play in Signaling Games," *The American Economic Review*, June 2005, *95* (3), 477–509.

De Roos, Nicolas and Yianis Sarafidis, "Desision Making under Risk in Deal or No Deal," *Journal of Applied Econometrics*, September 2009. Published online in advance of print.

Dohmen, Thomas, Armin Falk, David Huffman, Felix Marklein, and Uwe Sunde, "Biased probability judgment: Evidence of incidence and relationship to economic outcomes from a representative sample," *Journal of Economic Behavior & Organization*, December 2009, *72* (3), 903–915.

_ , _ , _ , **Uwe Sunde, Jürgen Schupp, and Gert Wagner**, "Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences," Technical Report, Maastricht: ROA, Research Centre for Education and the Labour Market 2009.

Donkers, Bas, Bertrand Melenberg, and Arthur Van Soest, "Estimating Risk Attitudes using Lotteries: A Large Sample Approach," *Journal of Risk and Uncertainty*, 2001, *22*, 165–195.

Fehr-Duda, Helga, Adrian Bruhin, Thomas Epper, and Renate Schubert, "Rationality on the rise: Why relative risk aversion increases with stake size," *Journal of Risk and Uncertainty*, April 2010, *40* (2), 147–180.

_ , **Manuele de Gennaro, and Renate Schubert**, "Gender, Financial Risk, and Probability Weights," *Theory and Decision*, 2006, *60*, 283–313.

_ , **Thomas Epper, Adrian Bruhin, and Renate Schubert**, "Risk and rationality: The effects of mood and decision rules on probability weighting," *Journal of Economic Behavior & Organization*, 2011, *78* (1-2), 14 – 24.

Fullenkamp, Connel, Rafael Tenorio, and Robert Battalio, "Assessing Individual Risk Attitudes Using Field Data From Lottery Games," *The Review of Economics and Statistics*, 01 2003, *85* (1), 218–226.

Gee, Christopher, "Risky Choice and Type-Uncertainty in "Deal or No Deal?"," Cambridge Working Papers in Economics 0758, Faculty of Economics, University of Cambridge November 2007.

Gertner, Robert, "Game Shows and Economic Behavior: Risk-Taking on "Card Sharks"," *The Quarterly Journal of Economics*, May 1993, *108* (2), 507–521.

Hall, Robert E., "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 1988, *96* (2), 339–57.

Harrison, Glenn W. and John A. List, "Field Experiments," *Journal of Economic Literature*, 2004, *42* (4), 1009–1055.

Hartley, Roger, Gauthier Lanot, and Ian Walker, "Who really wants to be a millionaire? Estimates of risk aversion from gameshow data," Technical Report, University of Warwick, Department of Economics 2006.

Hersch, Philip L. and Gerald S. McDougall, "Decision Making under Uncertainty When the Stakes Are High: Evidence from a Lottery Game Show," *Southern Economic Journal*, July 1997, *64* (1), 75–84.

Holt, Charles A. and Susan K. Laury, "Risk Aversion and Incentive Effects," *The American Economic Review*, December 2002, *92* (5), 1644–1655.

Jianakoplos, Nancy and Alexandra Bernasek, "Are Women more risk averse?," *Economic Inquiry*, 1998, *36* (4), 620–630.

Kahneman, Daniel and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, March 1979, *47* (2), 263–291.

Knight, Frank H., *Risk, Uncertainty, and Profit*, Library of Economics and Liberty, 1921. Retrieved May 17, 2011 from http://www.econlib.org/library/Knight/knRUP.html.

List, John A., "Friend or Foe? A Natural Experiment of the Prisoner's Dilemma," *The Review of Economics and Statistics*, 08 2006, *88* (3), 463–471.

Matsen, Egil and Bjarne Strøm, "Dominated choices in a simple game with large stakes," *Experimental Economics*, 2010, *13*, 99–119.

Metrick, Andrew, "A Natural Experiment in "Jeopardy!"," *The American Economic Review*, March 1995, *85* (1), 240–253.

Post, Thierry, Martijn J. van den Assem, Guido Baltussen, and Richard H. Thaler, "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show," *The American Economic Review*, March 2008, *98* (1), 38–71.

Prather, Larry J. and Karen L. Middleton, "Are N+1 heads better than one?: The case of mutual fund managers," *Journal of Economic Behavior & Organization*, January 2002, *47* (1), 103–120.

Rabin, Matthew, "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, September 2000, *68* (5), 1281–1292.

Rockenbach, Bettina, Abdolkarim Sadrieh, and Barbara Mathauschek, "Teams take the better risks," *Journal of Economic Behavior & Organization*, July 2007, *63* (3), 412–422.

Appendix: Tables

TABLE A: Variable definitions

Variable name	Definition/examples
Stake (/100)	Acquired winnings at the end of the "Quiz Taxi" ride in EUR, divided by 100 (when losing the third life, or when reaching the master question)
Stake per contestant (/100)	Acquired winnings at the end of the "Quiz Taxi" ride in EUR per group member, divided by 100 (when losing the third life, or when reaching the master question)
Distance	Distance from start to final destination in kilometers
Number of questions	Number of questions played
Lives left	Lives left at the end of the "Quiz Taxi" ride (0 - 3)
Wild card left	Telephone or passersby wild card not used during "Quiz Taxi" ride (dummy variable)
Share of right answers (in %)	Correct answers divided by number of questions in percent
Share of known answers (in %)	Known answers (no use of wild cards, no obvious guesses) divided by number of questions in percent
Streak	Number of consecutive correct answers directly before master question
2 passengers	Two contestants (dummy variable)
3 passengers	Three contestants (dummy variable)
Females only	Only female contestants (dummy variable)
Males only	Only male contestants (dummy variable)
Females and males	Female and male contestants (dummy variable)
Young contestants (under 30)	All contestants younger than 30 (dummy variable)
Middle age contestants (30 - 50)	All contestants between 30 and 50 (dummy variable)
Old contestants (above 50)	All contestants older than 50 (dummy variable)
Contestants of different age	Contestants of different age (dummy variable)
Migration background	At least one groups member appears to have a migration background, based on language skills and appearance (dummy variable)
DVD episode	Run comes from DVD, not from internet portal "maxdome.de" (dummy variable)
Discussion time	Length of discussion before the final decision to play the master question in seconds
Majority pro master	More contestants initially want to play master question (dummy variable)
Majority contra master	More contestants initially do not want to play master question (dummy variable)
Number of topics discussed	Number of arguments exchanged
Number of topics pro	Number of arguments that imply that master question should be played
Number of topics contra	Number of arguments that imply that master question should not be played
Humility/modesty	"Modesty is a virtue!" "Greed is not good¡' (dummy variable)
A lot of money	"That's good money for us." "That would be a whole monthly salary." (w.r.t. stake only, dummy variable)
Nothing to lose	"Well, we entered with nothing, so we can't really lose anything!" (dummy variable)
Provocation/encouragement by host	"You cowards!" "You did so well so far." (dummy variable)
Easy to answer master question	"It's not that hard, anyway." "It's not more difficult than the others." (only if mentioned by contestants, dummy variable)
Difficult to answer master question	"We've been so lucky so far." "We were not that good so far." (only if mentioned by contestants, dummy variable)

Source: Own transcripts

TABLE B: Variable means by topics discussed

	Humble		A lot of money		Nothing to lose		Provocation		Easy		Difficult	
	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Stake (/100)	8.10	8.26	8.06	8.44	8.26	7.18	8.20	7.96	8.21	7.13	8.13	8.06
Stake per contestant (/100)	3.78	3.72	3.75	3.85	3.83	3.36	3.85	3.60	3.82	3.21	3.76	3.79
Share of correct answers (in %)	84.43	84.10	83.85	87.08	84.44	83.98	83.65	85.88	84.36	84.61	84.25	84.94
Share of known answers (in %)	65.18	66.57	65.05	67.06	65.01	67.85	65.07	66.03	65.85	60.14	65.54	64.70
Discussion time (in seconds)	26.77	30.14	25.78	34.75	25.39	39.74	20.61	40.81	25.09	51.42	24.17	40.85
3 passengers	0.18	0.24	0.17	0.25	0.19	0.16	0.16	0.23	0.18	0.25	0.19	0.15
Females only	0.25	0.19	0.25	0.17	0.24	0.21	0.29	0.15	0.24	0.25	0.24	0.26
Males only	0.31	0.33	0.30	0.42	0.29	0.47	0.26	0.44	0.31	0.33	0.33	0.26
Observations	125	21	122	24	127	19	98	48	134	12	119	27

Source: Own calculations

TABLE C: Robustness checks: different estimation models
(Average marginal effects from specified regressions)

Dependent variable:	**Probit selection**				
Play master question (1=yes, 0=no)	**Probit**	**1st stage**	**2nd stage**	**OLS**	**Logit**
Stake (/100)	-0.044***	0.109***	-0.045*	-0.044***	-0.043***
	(0.016)	(0.007)	(0.027)	(0.016)	(0.016)
Share of correct answers (in %)	0.013**	–	–	0.014*	0.012*
	(0.006)			(0.007)	(0.006)
Wild card left	0.075	0.019	0.083	0.083	0.079
	(0.093)	(0.062)	(0.098)	(0.107)	(0.095)
3 passengers	0.071	-0.121*	0.044	0.078	0.077
	(0.106)	(0.063)	(0.122)	(0.114)	(0.113)
Females only	-0.138	-0.061	-0.147	-0.138	-0.141
	(0.095)	(0.062)	(0.103)	(0.094)	(0.099)
Males only	-0.033	-0.023	-0.037	-0.037	-0.036
	(0.090)	(0.057)	(0.095)	(0.097)	(0.093)
DVD episode	0.270**	0.061	0.291**	0.301*	0.267**
	(0.129)	(0.086)	(0.145)	(0.153)	(0.133)
Distance	–	-0.085***	–	–	–
		(0.017)			
Pseudo/Adjusted R2	0.09			0.07	0.09
Model p-value	0.027		0.41	0.014	0.049
Observations	146		256	146	146

Notes: Table reports average marginal effects from specified regressions, robust standard errors in parentheses. Selection model: Heckman maximum-likelihood estimator; first stage instrument for reaching the master question: distance to final destination. Significance levels: *10% **5% ***1%
Source: Own calculations

TABLE D: Robustness checks: different sets of covariates
(Average marginal effects from probit regressions; all groups)

Dependent variable: Play master question (1=yes, 0=no)	**Additional Covariates**	**Stake per contestant**	**Log of stake**	**Interaction stake * % correct**	**Share of known answers**	**Streak of correct answers**	**Log of discussion time**
Stake (/100)	-0.036**	–	–	-0.041**	-0.035**	-0.031**	-0.037**
	(0.016)			(0.016)	(0.017)	(0.015)	(0.016)
Stake per contestant (/100)	–	-0.070**	–	–	–	–	–
		(0.034)					
Log of stake	–	–	-0.040**	–	–	–	–
			(0.018)				
Share of correct answers (in %)	0.010	0.010	0.011	0.013**	–	–	0.011*
	(0.007)	(0.007)	(0.007)	(0.007)			(0.007)
Share of known answers (in %)	–	–	–	–	0.005	–	–
					(0.004)		
Streak of correct answers	–	–	–	–	–	0.017	–
						(0.014)	
Wild card left	0.065	0.068	0.067	0.063	0.027	0.066	0.072
	(0.095)	(0.095)	(0.095)	(0.094)	(0.097)	(0.094)	(0.094)
3 passengers	0.101	-0.008	0.104	0.121	0.092	0.076	0.087
	(0.102)	(0.098)	(0.102)	(0.101)	(0.104)	(0.101)	(0.103)
Females only	-0.110	-0.112	-0.106	-0.103	-0.110	-0.118	-0.107
	(0.093)	(0.093)	(0.093)	(0.094)	(0.094)	(0.094)	(0.095)
Males only	-0.021	-0.023	-0.019	-0.014	-0.014	-0.010	-0.031
	(0.088)	(0.088)	(0.088)	(0.088)	(0.087)	(0.088)	(0.088)
DVD episode	0.170	0.177	0.167	0.144	0.184	0.185	0.190
	(0.135)	(0.134)	(0.134)	(0.134)	(0.136)	(0.136)	(0.131)
Discussion time (in seconds)	0.005***	0.005***	0.005***	0.005***	0.006***	0.005***	–
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	
Log of discussion time	–	–	–	–	–	–	0.007**
							(0.003)
Young contestants (under 30)	-0.060	-0.064	-0.062	-0.059	-0.032	-0.050	-0.059
	(0.081)	(0.081)	(0.081)	(0.081)	(0.080)	(0.080)	(0.081)
Old contestants (above 50)	-0.043	-0.048	-0.061	-0.067	0.020	-0.040	-0.061
	(0.155)	(0.156)	(0.152)	(0.152)	(0.162)	(0.156)	(0.158)
Migration background	0.029	0.025	0.022	0.009	0.027	0.018	0.028
	(0.109)	(0.110)	(0.110)	(0.111)	(0.112)	(0.110)	(0.110)
Pseudo R2	0.13	0.13	0.13	0.14	0.13	0.13	0.12
Model p-value	0.042	0.046	0.045	0.025	0.043	0.033	0.038
Observations	146	146	146	146	146	146	146

Notes: Table reports average marginal effects from probit regressions, robust standard errors in parentheses. Interaction term included in regression of column 4, corresponding marginal effect not reported. Significance levels: *10% **5% ***1%
Source: Own calculations

TABLE E: Robustness checks: different sets of covariates
(Average marginal effects from probit regressions; undecided groups only)

Dependent variable: **Play master question (1=yes, 0=no)**	**Additional Covariates**	**Stake per contestant**	**Log of stake**	**Interaction stake * % correct**	**Share of known answers**	**Streak of correct answers**	**Log of discussion time**
Stake (/100)	-0.037	–	–	-0.033	-0.017	-0.011	-0.036
	(0.028)			(0.028)	(0.025)	(0.025)	(0.028)
Stake per contestant (/100)	–	-0.074	–	–	–	–	–
		(0.060)					
Log of stake	–	–	-0.028	–	–	–	–
			(0.031)				
Share of correct answers (in %)	0.030**	0.029**	0.028**	0.027**	–	–	0.031**
	(0.013)	(0.013)	(0.013)	(0.013)			(0.013)
Share of known answers (in %)	–	–	–	–	0.011**	–	–
					(0.005)		
Streak of correct answers	–	–	–	–	–	0.037*	–
						(0.020)	
Wild card left	-0.071	-0.068	-0.071	-0.065	-0.115	-0.062	-0.084
	(0.117)	(0.118)	(0.118)	(0.117)	(0.130)	(0.128)	(0.115)
3 passengers	0.112	0.009	0.103	0.098	0.082	0.047	0.099
	(0.120)	(0.128)	(0.123)	(0.122)	(0.130)	(0.124)	(0.119)
Females only	-0.302**	-0.305**	-0.304**	-0.310**	-0.254*	-0.286*	-0.299**
	(0.142)	(0.143)	(0.144)	(0.141)	(0.134)	(0.149)	(0.143)
Males only	-0.106	-0.110	-0.103	-0.126	-0.096	-0.059	-0.106
	(0.116)	(0.115)	(0.117)	(0.118)	(0.127)	(0.127)	(0.115)
DVD episode	-0.009	-0.001	-0.005	0.001	0.006	0.016	-0.003
	(0.169)	(0.168)	(0.168)	(0.170)	(0.175)	(0.165)	(0.165)
Discussion time (in seconds)	0.006**	0.006**	0.006**	0.006**	0.007**	0.007**	–
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	
Log of discussion time	–	–	–	–	–	–	0.010**
							(0.004)
Young contestants (under 30)	-0.121	-0.130	-0.128	-0.129	-0.139	-0.098	-0.120
	(0.101)	(0.101)	(0.103)	(0.101)	(0.113)	(0.113)	(0.101)
Old contestants (above 50)	0.096	0.090	0.052	0.200	0.022	-0.038	0.081
	(0.267)	(0.267)	(0.256)	(0.287)	(0.266)	(0.295)	(0.259)
Migration background	0.211	0.197	0.192	0.278**	0.206	0.125	0.229*
	(0.135)	(0.134)	(0.133)	(0.130)	(0.142)	(0.156)	(0.139)
Pseudo R2	0.23	0.23	0.22	0.24	0.17	0.17	0.22
Model p-value	0.128	0.126	0.124	0.078	0.161	0.301	0.119
Observations	65	65	65	65	65	65	65

Notes: Table reports average marginal effects from probit regressions, robust standard errors in parentheses. Interaction term included in regression of column 4, corresponding marginal effect not reported. Significance levels: *10% **5% ***1%
Source: Own calculations