Gender differences when subjective probabilities affect risky decisions: an analysis from the television game show Cash Cab

Matthew R. Kelley & Robert J. Lemke

Theory and Decision
An International Journal for Multidisciplinary Advances in Decision Science

ISSN 0040-5833
Volume 78
Number 1

Theory Decis (2015) 78:153-170
DOI 10.1007/s11238-013-9389-9

Springer

Theory Dec. (2015) 78:153–170
DOI 10.1007/s11238-013-9389-9

Gender differences when subjective probabilities affect risky decisions: an analysis from the television game show *Cash Cab*

Matthew R. Kelley · Robert J. Lemke

Published online: 11 July 2013

Abstract This study uses the television show *Cash Cab* as a natural experiment to investigate gender differences in decision making under uncertainty. As expected, men are much more likely to accept the end-of-game gamble than are women, but men and women appear to weigh performance variables differently when relying on subjective probabilities. At best men base their risky decisions on general aspects of their previous "good" play (not all of which is relevant at the time the decision is made) and at worst fail to condition their risky decisions on any of the relevant information available to them. In sharp contrast, women appear to consider all of the information available to them, including previous "poor" play as well as their most recent confident "good" play, which, by design, is likely the most relevant information to consider.

Keywords Subjective probabilities · Decision making under uncertainty · Female/male decision making · *Cash Cab*

JEL Classification D81 · C93 · L83 · J16

1 Introduction

The risk literature has long been interested in how decisions are made under uncertainty, and how those decisions differ across the sexes. In the present study, the tele-

M. R. Kelley
Department of Psychology, Lake Forest College, 555 N. Sheridan Road, Lake Forest, IL 60045, USA

R. J. Lemke (✉)
Department of Economics, Lake Forest College, 555 N. Sheridan Road, Lake Forest, IL 60045, USA
e-mail: lemke@lakeforest.edu

vision show *Cash Cab* is used as a natural experiment to investigate decision making under uncertainty when the probability of success is unknown and the decision maker has an opportunity to update her subjective probability of success. To our knowledge, this paper is the first to explore subjective updating within the context of a game show.

The literature largely agrees that women are more risk averse than men (e.g., for surveys, see Byrnes et al. 1999, Eckel and Grossman 2005). Booij et al. (2010), Fehr-Duda et al. (2006), and Jianakoplos and Bernasek (1998) find women to be more risk averse than men when making financial decisions. Cohen and Einav (2007) demonstrate similar gender differences in risk preferences when looking at choices of car insurance deductible amounts, and Hersch (1996) shows that women are less likely than men to engage in risky social behavior such as smoking and not wearing a seat belt. Two reasons are commonly given for this difference in risk aversion across the sexes: (1) men and women value payoffs differently, and (2) men and women differ in how they evaluate or process probabilities. Of these two, there is general consensus that the second is more important than the first. Moreover, Bruhin et al. (2010) show that women's weighting curves of probabilities are more nonlinear than men's (e.g., women seem to over-estimate small probabilities), which results in differing assessments of the expected value of a gamble. Borghans et al. (2009), however, argue that the difference in probability assessment across sexes diminishes as the uncertainty of the gamble increases (called ambiguity aversion by Ellsberg 1961).

Beginning with Gertner (1993), game shows have been used to investigate decision making under uncertainty. Game shows offer two main benefits: payouts are substantially larger than what can usually be offered in a laboratory environment, and most game shows involve straightforward, simple probabilities. *Cash Cab*, however, is different. While answering a series of general knowledge questions (ungrouped by category), contestants on *Cash Cab* can learn about their ability to answer the types of trivia questions asked on the show. At the end of each game, a successful contestant is presented with a gamble: take her winnings (which average more than \$850) and leave the show, or risk her winnings double-or-nothing on one last trivia question (topic not indicated). When presented with this gamble, a contestant has the opportunity to assess her subjective probability of being able to answer the question correctly, as she decides to either accept or reject the final gamble.

To preview, the results of the present study show clear differences across the sexes in their willingness to risk their winnings on the double-or-nothing gamble. In general, men are much more likely than women to accept the gamble. We also find that contestants appear to update their subjective probabilities based on the confidence of their previous answers rather than relying solely on whether their previous answers were correct. Most importantly, however, men and women seem to focus on different aspects of their game play when considering the final gamble. Female contestants are less likely to accept the gamble if they were previously asked a question for which they had no confidence in their answer while male contestants show no such tendency. Furthermore, female contestants appear to consider how confident they were in correctly answering their more recent questions (which according to *Cash Cab* are the hardest), whereas male contestants consider their confidence in all of their previous questions, including the easier ones. Given the structure of *Cash Cab*, therefore, women appear to condition their decision making on the most relevant data available to them, whereas men tend to consider less relevant information when updating. These results run counter to previous findings that women are more likely than men to form biased assessments of probabilities.

The paper is organized as follows. In the next section, we describe the television game show *Cash Cab*. Following that, we detail the data collection methodology and report the descriptive statistics. Section 4 discusses how decisions made on trivia game shows potentially rely on subjective probabilities and the updating of those probabilities. The empirical results are presented in Sect. 5 and the general discussion is Sect. 6.

2 Cash cab

Cash Cab is a television game show that aired on the Discovery Network in the United States from 2005 to 2012. The show consists of an unidentifiable taxi cab driven in Manhattan by the host of the show, Ben Bailey. After entering the cab and giving a destination, the host identifies himself and informs the passengers of the game. The passengers are then given the opportunity to leave the cab and not play or to start the game. All rides originate and terminate in Manhattan. The game consists of the contestants (acting as a single team) winning money by answering a series of trivia questions asked by the host. The game ends either when the contestants answer three questions wrong (i.e., they receive three strikes) and no money is won, or when the cab reaches the contestant's destination in which case the contestants are entitled to all the money they have accumulated. While playing the game, contestants have two shout-outs they can use at any time—a "mobile shout-out" in which they can call anyone they know for help on a question and a "street shout-out" in which the cab pulls to the curb and the contestants invite a passerby over to the cab for help on the question.

Each of the first four questions of a game is worth a fixed initial value, $25, $50, or $100, depending on the production season and whether the passengers have been randomly selected for a Cash Cab Double Ride.[1] The values of questions 5 to 8 are double the initial value, while all questions after the 8th are worth quadruple the initial value. The contestants are also told that the questions get more difficult with each step up in dollar value.

Although not by design, *Cash Cab*'s somewhat random assignment of dollar values to rides is convenient for our analysis as winnings are not closely tied to the contestant's performance. For example, completing one's ride with $800 could be achieved by answering 13 of 13 questions correctly without the help of any shout-outs on a $25 ride or by knowing the answers to just 3 of 7 questions on a $100 ride in which both shout-outs were successful in providing 2 additional correct answers.

Finally, contestants who reach their destination before receiving three strikes, termed a "completed ride," are presented with a gamble. They can take their accumulated winnings and exit the cab, or they can risk all of their winnings and go

[1] The initial value was set at $25 for the first two production seasons, and double rides did not occur. After that, the initial value was set at $50, with an initial value of $100 for Double Rides. In our sample, from season three on, 13.5 percent of rides have been Double Rides.

double-or-nothing on a "video bonus question (VBQ)." The host informs the contestants that if they risk their money on the VBQ, they will be shown a short video and asked a single question that is more difficult than the previously asked questions.[2] Contestants cannot use any remaining shout-outs to help answer the VBQ. If the contestants answer the VBQ correctly, their winnings are doubled. If they fail to answer the VBQ correctly, they lose all accumulated winnings and exit the cab at their destination having won nothing except a free cab ride. Throughout the paper, we refer to this portion of the game as the "VBQ gamble."

The economic analysis of this paper concerns the VBQ gamble as it provides a natural experiment regarding decision making under uncertainty. That is, contestants who successfully reached their destination had, roughly 15 to 30 min earlier, no idea that they would be offered a gamble that asked them to choose between (a) a safe option in which they are guaranteed some money and (b) a risky double-or-nothing gamble that depends on answering a single trivia question correctly. Being presented with this gamble, therefore, allows us to investigate decision making under uncertainty.

It is important to recognize the features of *Cash Cab* that are particularly attractive for this type of study. First, potential winnings are non-trivial. Lab and classroom experiments are often criticized for not offering gambles with substantial sums to be won. Game shows became a popular source of data in part because they offer the potential for significant winnings. Although *Cash Cab* does not offer monetary amounts at the same level of *Jeopardy* or *Deal or No Deal*, it still offers meaningful sums of money (average of $856 to risk on the VBQ)—anecdotally, when discussing the VBQ gamble, many contestants on *Cash Cab* mention that their winnings will cover rent or a credit card balance, so in the eyes of many *Cash Cab* contestants, the money is important. Second, unlike other game shows, contestants do not apply to be on *Cash Cab*. While riders on *Cash Cab* are not a random slice of society, as one must be hailing a cab in Manhattan and going someplace else in Manhattan, they likely represent a broader group of people than appear on any other game show. Inferring results from *Cash Cab* contestants to the general population, therefore, probably rests on safer ground than doing so with other game shows. Finally, given that *Cash Cab* is a trivia show and the topic of the VBQ in unknown, the contestants have the opportunity to update their subjective probability of correctly answering the VBQ by considering their previous answers and the confidence with which they gave those answers.

3 Data collection and descriptive statistics

Data were collected from all 196 episodes of *Cash Cab* that aired during the 26 weeks from 15 May 2011 to 12 Nov 2011. The only episodes omitted from the analysis were eight episodes from the first season. These eight shows were played slightly differently than the rest and, in particular, the host was more helpful than in later shows by reminding the contestants of the rules of the game and encouraging the use of shout-outs. The data were collected by having both authors watch each show in its

[2] Empirically, 71 percent of the completed rides that accept the gamble answer the VBQ correctly, whereas this same group answered 85 percent of its previous questions correctly.

entirety. Two types of data were collected—objective (e.g., response accuracy) and subjective (e.g., response confidence). Any discrepancies in the objective data were corrected by returning to the recording of the show. When there was a discrepancy in subjective data, consensus was achieved and the observation was flagged. All except two shows included three "rides" (i.e., games) with each ride numbering from one to four contestants.[3]

At the start of each ride, the sex and race of each contestant, the distance of the ride measured in city blocks, and whether the ride took place in the daytime or nighttime were recorded. An estimate of each contestant's age was also made, with each contestant being classified as being under the age of 18, between 18 and 30, between 31 and 50, between 51 and 65, or being older than 65. (Although each ride in the *Cash Cab* can include up to four riders and of both sexes, to facilitate the writing of the results we refer to a single female contestant competing on the show throughout the paper.)

During the game itself, data were recorded on each question as to whether the contestant provided an accurate response and whether a shout-out was used. Additionally, the contestant's confidence in each answer was rated by the authors as falling into one of three categories—highly confident, somewhat confident, and not at all confident. Confidence was assessed by considering contestants' tone and body language immediately prior to and as they answered the question, and not by whether the contestants actually answered a question correctly. On rides with two or more players, confidence was assessed for the entire group in that the maximum confidence exuded by any one of the contestants on a particular question was recorded.[4]

The 196 episodes of *Cash Cab* yielded data on 586 rides. Of these, 379 (64.7 %) resulted in the contestant completing the ride and being presented with the VBQ gamble. Of these, 159 (42 %) took the gamble and risked their winnings, while 220 (58 %) declined the gamble. Table 1 presents the summary statistics for all of the variables used in the analysis. The first four columns report the minimum and maximum values along with the mean and standard deviation for each variable for the entire sample of 586 rides. The remaining six columns report the conditional mean and standard deviation for three samples of interest—all completed rides, all completed rides that risked their winnings on the VBQ, and all completed rides that chose to not risk their winnings. The first six rows of Table 1 are self-explanatory. On average, each game included almost 2.4 riders, a destination 37 blocks from where the cab was hailed, an average contestant age of 35.6, 47.6 percent of riders were female, and 85 percent of riders were white.

The next seven rows of Table 1 provide information on the performance of contestants. On average, rides use 0.927 shout-outs, but this falls to 0.834 for completed rides and to just 0.761 for completed rides that accept the gamble. Completed rides that did not accept the gamble averaged using 0.886 shout-outs.

[3] Babies were not included in the number of contestants, but children who could participate were included (and occasionally children provided correct answers). One ride included 5 contestants, with three of the five being children. We classified this as a ride of 4 contestants.

[4] The confidence judgment was recorded by each author prior to hearing whether the answer was correct, and their independent assessment of confidence matched on over 93 percent of all questions.

Table 1 Descriptive statistics

	All rides (N = 586)				Completed rides (N = 379)		Risk-takers (N = 159)		Non-risk-takers (N = 220)	
	Min.	Max.	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
Number of riders	1	4	2.374	0.886	2.454	0.857	2.642	0.866	2.318	0.827
Distance of ride in city blocks	6	83	37.0	8.2	36.2	8.0	35.6	7.8	36.7	8.1
Daytime ride	0	1	0.589	0.492	0.594	0.492	0.541	0.500	0.632	0.483
Average age of riders	26	70	35.6	9.7	35.9	9.5	33.9	8.5	37.3	10.0
Percent of riders who are female	0	100	47.6	36.6	45.1	35.36	41.1	33.7	48.0	36.3
Percent of riders who are white	0	100	85.0	30.4	86.5	28.6	88.0	25.5	85.5	30.6
Number of shout-outs used	0	2	0.927	0.746	0.834	0.728	0.761	0.724	0.886	0.728
Number of questions asked	3	20	9.2	2.6	10.0	2.4	9.8	2.2	10.1	2.5
Number of correct answers	0	19	7.2	3.0	8.6	2.4	8.4	2.3	8.7	2.6
Number highly confident correct answers	0	16	5.7	2.9	6.7	2.5	6.9	2.4	6.6	2.7
Number of strikes	0	3	2.0	1.0	1.4	0.7	1.4	0.7	1.4	0.7
Number answers with no confidence	0	5	1.6	1.1	1.3	1.0	1.1	0.9	1.4	1.0
Streak of highly confident correct answers to end the game	0	11	0.8	1.6	1.2	1.9	1.4	2.1	1.0	1.7
Successfully reached destination	0	1	0.647	0.478	1	0	1	0	1	0
Money winnings before gamble	$0	$3,100	$554	$545	$856	$448	$790	$413	$904	$467
Risked winnings on video bonus	0	1	n.a.	n.a.	0.419	0.494	1	0	0	0
Answered video bonus correctly	0	1	n.a.	n.a.	n.a.	n.a.	0.711	0.455	n.a.	n.a.
Money winnings at end of game	$0	$6,200	$656	$787	$1,014	$771	$1,166	$1,039	$904	$467

Source Author's calculations from self-collected data from *Cash Cab* episodes that aired on the Discovery network from 15 May 2011 to 12 Nov 2011

Looking at all completed rides, the typical ride was presented with 10 questions, 8.6 of which were answered correctly (possibly with the assistance of shout-outs).

An important aspect of the analysis concerns the contestant's confidence (high, some, or none) when giving answers. For all rides, contestants averaged giving 5.7 answers with high confidence that were also correct. This increases to 6.7 questions for completed rides and increases further (decreases) to 6.9 (6.6) for contestants that accepted (rejected) the gamble.

A similar pattern exists for the number of strikes received during the game and the number of answers given with no confidence. It is important to recognize the difference in these variables. The number of strikes is simply the number of strikes a contestant has at the end of the ride (i.e., the number of questions answered wrong). In contrast, the number of answers given with no confidence is the total number of answers given with no confidence whether or not the contestant received a strike, including all answers given after receiving help from a shout-out.

We also count the number of correct answers given with high confidence to end the game. This variable is important because it allows the analysis to identify how the quality of a contestant's most recent play affects her decision making. Because questions get harder over time, and because the game ends on a third strike, long streaks of highly confident correct answers to finish the game are rare. One ride finished on a streak of 11 such answers, but the average for the entire sample was just 0.8. For completed rides this increases to 1.2. Of completed rides that risk their winnings on the VBQ, the average ending streak is 1.4 highly confident and correct answers, whereas it is only 1.0 for contestants that do not risk their winnings.

The last five rows of Table 1 pertain to the end of the game and the VBQ gamble. Almost 65 percent of rides reach their destination (with average winnings of $856) and are presented with the VBQ. The raw data suggest that contestants are less likely to accept the gamble the more money that is at risk as average winnings prior to the gamble of those rides that accept the gamble is $790, whereas it is $904 for those rides that decline the gamble. Of all completed rides, 41.9 percent accept the gamble, and of those, 71.1 percent answer the VBQ correctly.

4 Trivia game shows and subjective probabilities

To set the stage for our empirical work, a brief review of the game show literature is warranted. A large literature examines play on a wide variety of game shows to assess risk parameters, including *Card Sharks* (Gertner 1993), *Deal or No Deal* (Brooks et al. 2009; De Roos and Sarafidis 2010; Deck et al. 2008; and Post et al. 2008), *Cash Cab* (Bliss et al. 2012; Keldenich and Klemm 2012), and on televised large-stake, state lottery games (Fullenkamp et al. 2003; Hersch and McDougall 1997). In all of these games except *Cash Cab*, the contestants (and the researcher) know the probability or expected value associated with every gamble presented in the game.[5] When these games are played in rounds, the probabilities can change, but the required updating is mathematically clear. Thus, contestants may need to update probabilities from round to round, but there is no role for *subjective probabilities* (or the updating of subjective probabilities).

[5] In each of these games, the probabilities are either obvious or fairly straightforward to calculate. Even in *Deal, No Deal,* in which there are 26 briefcases to keep track of, the show provides the contestant a visual aid to understand what options remain.

In contrast, trivia game shows are not associated with known probabilities. Skill and knowledge affect the likelihood of having a successful game. For trivia games, therefore, contestants are likely to continuously update their beliefs regarding their own chances of success as the game progresses. In this regard, the probability associated with the VBQ in *Cash Cab* is a subjective probability—the contestant makes a subjective guess as to her chance of answering the VBQ correctly and then accepts or rejects the VBQ gamble. Many factors potentially affect the contestant's ultimate decision whether to accept the VBQ, including risk attitudes, personal wealth, one's relationship to the other contestants, and the group's subjective probability of being able to correctly answer the VBQ, just to name a few.

The VBQ gamble of *Cash Cab*, therefore, may offer an opportunity to investigate updating of subjective probabilities. We are, of course, never in a position to know the contestant's subjective probability or to know precisely how the contestant has updated her beliefs. Still, the decision to accept or reject the VBQ likely incorporates, to some extent, the contestant's subjective probability—the greater a contestant's subjective probability of providing the correct answer, the more likely she is to gamble on the VBQ and vice-versa. Because we have detailed data about the contestant's performance throughout the game, we are in position to ask what particular aspects about the contestant's play are most important in making the VBQ decision, and therefore to learn which aspects might influence the updating of subjective probabilities.

If contestants use their game play to update their subjective probability of answering the VBQ correctly, then the subjective probability of answering the VBQ correctly should be higher for contestants that knew more correct answers during the course of the ride. Because we also collected confidence data, however, we can move beyond just looking at right and wrong answers. For instance, the contestant's confidence in her ability to answer a question on an unknown topic might become an important factor in the determination of her subjective probability. A contestant who confidently answered many questions would reasonably be more confident in her ability to answer the VBQ correctly than a contestant who lacked this confidence in her previous answers. Moreover, as the questions on *Cash Cab* get more difficult after the 4th and again after the 8th question, a contestant may be best to consider her accuracy and confidence on her later questions (i.e., her harder questions) more than she considers her performance on her initial (easier) questions.

5 Results

Although the focus of the paper concerns the effect of gender on decision making under uncertainty, we first consider how a variety of pre-game (Sect. 5.1) and in-game factors (Sect. 5.2) relate to the VBQ decision across the sample. These analyses will show that behavior on *Cash Cab* is consistent with well-established results in the literature and will also highlight factors of game play that are uniquely found on *Cash Cab*. We then provide a detailed examination of gender (Sect. 5.3) and show that all-female and all-male groups appear to be sensitive to different in-game factors when considering the VBQ, which might reflect differences in subjective updating across the sexes.

5.1 Pre-game factors that contribute to the VBQ decision

To begin, we estimate the VBQ decision as a function of only those factors observable at the start of the game (i.e., pre-game factors: gender, race, age, group size, and time of day). Model (1) in Table 2 provides these results. The dependent variable equals 1 if the contestant accepted the VBQ gamble and risked her winnings on a double-or-nothing bet and equals 0 if the contestant declined the VBQ gamble and instead took her current winnings and left the game. In the model, we include dummy variables for "half-female ride" and "all-male ride" compared against the omitted group of "all-female ride,"[6] the percent of riders who were white, the average age of all riders, dummy variables for one-, two-, and three-rider groups compared to four-rider groups, and whether the ride started during the daytime compared to a nighttime ride. All regression results (Tables 2, 3) are generated using a linear probability model (LPM) with robust standard errors. The benefit of using LPM estimation is that the coefficients are interpreted as percentage point changes in the decision to accept the gamble.[7]

Past research has shown that men are more likely to take risks than women (for a meta-analysis, see Byrnes et al. 1999), whites tend to perceive less risk than nonwhites (e.g., Finucane et al. 2000; Flynn et al. 1994), and the willingness to take risks decreases as people grow older (e.g., Barsky et al. 1997; Donkers et al. 2001; Jianakoplos and Bernasek 1998). Our results confirm each of these findings: (a) all-male rides are 17.09 percentage points more likely to take the gamble compared to all-female rides (p value = 0.020), (b) compared to an all-nonwhite ride, an all-white ride is expected to be 17 percentage points more likely to risk winnings on the VBQ, and (c) a completed ride is 6.3 percentage points less likely to take the gamble when the average age of the riders increases by 10 years.[8]

In addition, the presence of other contestants and the size of the group have been shown to influence risky decisions, as individuals tend to take more risks when in groups than when alone (e.g., Gardner and Steinberg 2005; Keldenich and Klemm

[6] Because of our interest in gender differences, we limit the sample to three types of rides: all-female rides ($n = 73$), half-female rides ($n = 119$) being either one female and one male or two females and two males, and all-male rides ($n = 108$). We remove 79 three-person, multi-sex rides from the analysis in order to better focus on the three gender groups for which we have large enough sample sizes for regression analysis. Future research might investigate how three-person rides make the VBQ decision and how having more men or more women matters in group decision making (Ertac and Gurdal 2012; Ronay and Kim 2006).

[7] Because of the limitations of LPM estimation, a data appendix containing all regression results under logit regression are available upon request. The results from the logit regressions are qualitatively unchanged from the LPM results reported throughout the paper.

[8] None of the estimated coefficients change significantly when the distance of the ride is included in the model, and the estimated coefficient on distance is statistically insignificant in all specifications. Therefore, none of the regression models include the distance of the ride.

Table 2 Regression results with subjective influences

	All questions			Questions 5 and after	
	(1)	(2)	(3)	(4)	(5)
Half of all riders are female	0.0286	0.0236	−0.0174	0.0261	0.0040
	0.0743	0.0743	0.0737	0.0747	0.0736
All riders are male	0.1709[b]	0.1620[b]	0.1204	0.1680[b]	0.1358[c]
	0.0733	0.0747	0.0757	0.0738	0.0741
Percent of riders white	0.0017[b]	0.0017[c]	0.0017[c]	0.0017[c]	0.0018[b]
	0.0009	0.0009	0.0009	0.0009	0.0009
Average age of riders	−0.0063[b]	−0.0058[b]	−0.0065[b]	−0.0060[b]	−0.0064[b]
	0.0027	0.0027	0.0027	0.0027	0.0027
One-person ride	−0.3738[a]	−0.3852[a]	−0.3604[a]	−0.3832[a]	−0.3687[a]
	0.1221	0.1240	0.1233	0.1239	0.1228
Two-person ride	−0.2064[b]	−0.2183[b]	−0.2132[b]	−0.2113[b]	−0.2160[b]
	0.0908	0.0922	0.0912	0.0929	0.0898
Three-person ride	−0.1646	−0.1724	−0.1802	−0.1696	−0.1713
	0.1192	0.1222	0.1205	0.1229	0.1204
Daytime ride	−0.0995[c]	−0.1118[c]	−0.1206[b]	−0.1110[c]	−0.1185[b]
	0.0565	0.0573	0.0567	0.0573	0.0568
Winnings before the VBQ (in $100)	−0.0624[a]	−0.0742[a]	−0.0817[a]	−0.0723[a]	−0.0822[a]
	0.0188	0.0212	0.0197	0.0211	0.0200
Squared winnings before VBQ (in $100)	0.0016[b]	0.0019[b]	0.0020[a]	0.0018[b]	0.0020[a]
	0.0007	0.0008	0.0007	0.0008	0.0008
Number of correct answers		0.0136		0.0141	
		0.0132		0.0141	
Number of strikes		−0.0408		−0.0171	
		0.0525		0.0497	
Number of correct answers to end the game		−0.0024		0.0009	
		0.0117		0.0117	
Number of highly confident correct answers			0.0262[b]		0.0312[b]
			0.0125		0.0145
Number of answers given with no confidence, including all shout-outs			−0.0625[b]		−0.0626[c]
			0.0287		0.0330
Number of highly confident correct answers to end the game			0.0027		0.0012
			0.0179		0.0184
Constant	1.0206[a]	1.0518[a]	1.1156[a]	1.0371[a]	1.1346[a]
	0.1666	0.1874	0.1671	0.1767	0.1657
R^2	0.1267	0.1326	0.1606	0.1305	0.1552

Notes The dependent variable equals 1 or 0 if the contestant accepted or rejected the gamble of the video bonus question respectively. Each model is estimated either using a linear probability regression model. Robust standard errors are reported beneath estimated coefficients. There are 300 observations. 1, 5, and 10 % significance levels are indicated by a, b, and c, respectively

Table 3 Regression results with gender interacted with subjective influences

	All questions	Questions 5 and after
	(1)	(2)
Number of highly confident correct answers		
All riders are female	0.0230	0.0263
	0.0189	0.0265
Half of all riders are female	0.0329[b]	0.0446[b]
	0.0153	0.0203
All riders are male	0.0271[c]	0.0314[c]
	0.0143	0.0184
Num. of answers given with no confidence, incl. shout-outs		
All riders are female	−0.0957[b]	−0.1049[b]
	0.0397	0.0475
Half of all riders are female	−0.0975[b]	−0.0959[b]
	0.0463	0.0486
All riders are male	−0.0093	−0.0216
	0.0445	0.0452
Number of highly confident correct answers to end the game		
All riders are female	0.0531[c]	0.0476
	0.0293	0.0326
Half of all riders are female	−0.0416	−0.0456
	0.0339	0.0305
All riders are male	−0.0004	0.0010
	0.0188	0.0188
R^2	0.1730	0.1599

Notes The dependent variable equals 1 or 0 if the contestant accepted or rejected the gamble of the video bonus question respectively. Each model is estimated using a linear probability regression model. Robust standard errors are reported beneath estimated coefficients. There are 300 observations. All regressions also include the rider's money winnings (and its square) before the VBQ, dummy variables for the number of riders, a dummy variable for a daytime ride, the average age of the riders, and the percent of riders who were white. 1, 5, and 10 % significance levels are indicated by a, b, and c, respectively

2012; Vinokur 1971).[9] Accordingly, we find that, compared to four-rider teams, rides with one, two, or three contestants are 37, 21, and 16 percentage points less likely to accept the gamble (the last estimate is not statistically significant). Finally, when watching episodes of *Cash Cab*, one realizes that the time of day matters in terms of the contestants' disposition—almost all nighttime rides are taking contestants to or from a social event, whereas daytime rides vary more as the contestants can be on a social

[9] The general consensus in the literature is that groups are more risk-loving than individuals, and our results for *Cash Cab* support that on the whole. Intuitively, however, the behavior on *Cash Cab* could have gone the other way as the contestants are friends, and more risk-loving friends could acquiesce to their risk-averse friends. Unfortunately, the data do not allow us to isolate these two confounding effects.

outing, with children, or in the cab because of a job. Although barely statistically significant in model (1), contestants are almost 10 percentage points less likely to accept the gamble if their ride took place during the daytime compared to the nighttime.

5.2 In-game performance and confidence factors that contribute to the VBQ decision

Table 2 also contains variables that capture how well the contestant actually performed during the game (i.e., in-game performance factors: amount of winnings at risk, number of correct answers, number of strikes, number correct to end the game). One standard prediction of decision making under uncertainty is that the more money at stake, the less likely one is to take the gamble (Jianakoplos and Bernasek 1998; Post et al. 2008). Model (1) of Table 2 includes the amount of winnings at risk and its square.[10] The point estimates suggest that the probability of accepting the gamble declines as winnings increase up to having won $1,950. After this point the probability of accepting the gamble increases with winnings.[11]

With regard to the number of correct answers, strikes, and the ending correct streak, the results of model (2) in Table 2 are striking for their lack of explanatory power—not one of these variables is statistically significant.[12] Performance during the game, measured purely as right or wrong, is unrelated to the VBQ decision. That is, contestants on *Cash Cab* do not appear to rely on response accuracy when deciding whether to accept the gamble. At first glance, one might interpret these results as suggesting that contestants do not update their subjective probability of future success based on previous outcomes. However, the in-game confidence data, presented below, might suggest otherwise.

In model (3) of Table 2, the number of correct answers is replaced with the number of correct answers given with high confidence, the number of strikes is replaced with the number of questions answered with no confidence, and the number of correct answers to end the game is replaced with the number of highly confident, correct

[10] We do not know which multi-rider games include family units, for which total winnings is likely the better measure affecting decisions, or friendships, for which per-share winnings is likely the better measure. Bliss et al. (2012) argue that, even though per-share measures of winnings are what should matter theoretically, *Cash Cab* contestants make decisions based on total winnings. Our results are quantitatively unchanged if we include total winnings or winnings per contestant. Following Bliss et al. (2012), therefore, all of the results reported in the paper include total winnings.

[11] Only twelve observations have winnings in excess of $1,950 when the VBQ is offered, and only three of these accepted the gamble. However, the ride with the greatest winnings ($3,100) accepted the VBQ gamble (and went on to answer it correctly).

[12] The empirical model includes the number of questions rather than the percent of questions, because of the shortness of the game. Over 70 percent of all rides encounter between seven to eleven questions. Thus, one question is roughly ten percent for most rides. Moreover, the idea that good or bad play may later enter into the VBQ decision in such a short game would seem to be an additive, not percentage, effect. The model also does not control for the number of questions asked. In model (2) of Table (3), and again in model (4), the number of total questions asked is a linear combination of the number of questions answered correctly and the number of strikes. In the other models, the variables are highly co-linear. The number of questions was never statistically significant when it was included in any of the models, so this variable is omitted from Tables 2 and 3. A data appendix of results is available upon request that includes performance variables as percentages and that includes the number of questions asked of each set of contestants.

answers to end the game. Replacing the number of strikes variable with the number of answers given with no confidence is not a minor change. Although it rarely happens, some contestants provide a lucky guess in such situations. More importantly, however, because *Cash Cab* allows each team up to two shout-outs, the raw number of strikes does not capture the number of times the contestant had no idea of the answer. The number of answers given with no confidence, therefore, includes every question for which a shout-out was used.

Although the end-of-game streak remains statistically insignificant (more on this in the next subsection), the first two confidence variables are statistically significant. The VBQ is 2.6 percentage points more likely to be accepted for each additional correct answer given with high confidence and is 6.25 percentage points less likely to be accepted for each additional answer given with no confidence. These results are consistent with the notion that contestants rely on their subjective probability of future success by considering relevant game play, specifically in the form of their in-game confidence, when considering the VBQ decision.

Because contestants on *Cash Cab* are told that the questions get harder after the 4th question (and again after the 8th) and are told that the VBQ is harder still, contestants should place more weight on these harder questions when contemplating the VBQ. To allow for this possibility, columns (4) and (5) of Table 2 report the estimated coefficients from repeating columns (2) and (3) but rather than measuring the in-game accuracy and confidence variables for the entire game, they are measured only from question 5 on.[13] The results in columns (4) and (5) closely match the previous ones—accuracy is not related to the decision, but confidence in answers is strongly related to the decision to accept or reject the gamble.[14]

5.3 Gender analysis of in-game performance and confidence factors

Table 3 includes estimated coefficients from two linear probability regressions. Although omitted from Table 3 for space considerations, both regressions include all of the pre-game variables that were included in Table 2. Model (1) of Table 3 is an extension of model (3) in Table 2 in that all three gender categories (all-female, half-female, and all-male rides) are interacted with the contestant's total number of highly confident correct answers, answers given with no confidence, and highly confident correct answers to end the game.

The results suggest there is a small but positive relationship between the number of questions answered correctly with high confidence and accepting the VBQ, particularly for all-male and half-female groups. In particular, each additional highly confident correct answer has no statistically significant relationship with the decision

[13] Measuring the variables from question 9 on is less useful as most rides do not receive more than 9 questions.

[14] Although the measures of accuracy and confidence in Tables 2 and 3 are correlated, multicolinearity does not appear to be a problem in any of the regressions as none of the variables are associated with a variance inflation factor above 3. If models (2) and (3) are combined into one, the variance inflation factors exceed 7 for most of the response variables. In this case, the response accuracy variables remain statistically insignificant, while the confidence variables keep their sign but are slightly less statistically significant.

to accept the VBQ gamble for all-female rides (point estimate of 0.0230 with a p value of 0.226), but is associated with a 3.29 percentage point increase in the chance of accepting the gamble (p value = 0.032) for rides with half women, and is associated with a 2.71 percentage point increase in the chance of accepting the gamble (p value = 0.058) for all-male rides. To be clear, the statistical significance on the estimated coefficients in column (1) is with regard to whether the estimated effect for the specific gender group is different from zero. A separate test that all three coefficients are equal fails to be rejected with a p value of 0.8670. A joint test that all three of these point estimates equal zero is associated with a p value of 0.1307. The results are similar in column (2) of Table 3 in which all of the variables are measured from question 5 on.

In stark contrast, also reported in model (1) of Table 3, a female presence in the cab (i.e., an all-female ride or half-female ride) is associated with being almost 10 percentage points less likely to accept the gamble for each additional question answered with no confidence. There is no such effect for all-male groups. Thus, women seem to allow their future decisions to be influenced by previous "bad play," whereas men do not consider this factor. This result is further supported in column (2).

Finally, the streak of highly confident, correct answers to end the game is interacted with gender group. Although the results are much less precisely estimated, they indicate that all-female groups take their ending streak into account, whereas the other groups do not. Extending this streak of answers by one is associated with all-female groups being 5.31 percentage points more likely to risk their winnings on the VBQ (p value = 0.071), whereas the relationship is not statistically significant for half-female or all-male rides. Thus, the data suggest that all-female groups are more influenced by "good play" to end the game than are mixed-gender or all-male groups.

6 Discussion

In contrast to much of the game show literature, we set out to provide a unique look into whether and how contestants might rely on subjective beliefs in a situation where objective probabilities do not exist and agents have the opportunity to learn as they play the game. The empirical results from *Cash Cab* indicate that a variety of pre-game and in-game factors influence the VBQ decision. Generally, the pre-game factors of gender, race, age, and group size are all consistent with the standard effects reported in the literature; males, whites, younger contestants, and larger groups are all more likely to risk winnings on the VBQ than are females, nonwhites, older contestants, and smaller groups, respectively. With regard to the in-game factors, although past correct performance is not related to the VBQ decision, both the amount of winnings and the number of highly confident correct responses are strongly related to the decision to accept the gamble. Furthermore, a fine-grained gender analysis of confidence shows that women are more influenced by both their "bad play" (no confidence) throughout the game and their most recent "confident good play" than are men, who in turn, appear to be influenced at best solely by their overall confident good play.

Despite the richness of the dataset, a variety of other potentially important pre-game factors simply are not available (e.g., socio-economic status, risk attitudes, trivia

Table 4 Means by gender class

	Completed rides	Risk-takers	Non-risk-takers
All-female rides			
Number of observations	73 (100 %)	22 (30 %)	51 (70 %)
Number of riders	2.16	2.09	2.20
Number of questions asked	9.47	9.50	9.45
Number of correct answers	8.03	8.23	7.94
Num. highly confident correct ans.	5.70	6.32	5.43
Num. answers with no confidence	1.62	1.23	1.78
Money winnings before gamble	$736	$706	$750
Half-female, half-male rides			
Number of observations	119 (100 %)	44 (37 %)	75 (63 %)
Number of riders	2.40	2.68	2.24
Number of questions asked	10.06	9.87	10.16
Number of correct answers	8.60	8.50	8.65
Num. highly confident correct ans.	6.81	6.93	6.73
Num. answers with no confidence	1.20	1.07	1.28
Money winnings before gamble	$788	$732	$820
All-male rides			
Number of observations	108 (100 %)	48 (44 %)	60 (56 %)
Number of riders	2.11	2.27	1.98
Number of questions asked	10.31	9.81	10.70
Number of correct answers	8.99	8.38	9.48
Num. highly confident correct ans.	7.15	6.83	7.40
Num. answers with no confidence	1.16	1.17	1.15
Money winnings before gamble	$931	$744	$1,080

Source Author's calculations from self-collected data from *Cash Cab* episodes that aired on the Discovery network from 15 May 2011 to 12 Nov 2011

ability, and confidence). Although we have no access to SES and pre-existing risk attitudes, the present data allow us to potentially infer whether there were inherent differences in trivia ability and confidence across the three gender groups. To this end, Table 4 lists the mean value of several variables for each of these three groups, both for completed rides (column 1) as well as for the subgroups of risk-takers (column 2) and non-risk-takers (column 3).

With respect to ability, performance measures indicate that the three gender groups do not differ in their abilities to answer trivia questions correctly. All-female and half-female rides both correctly answered 85 percent of all questions posed to them, while all-male rides correctly answered 87 percent of their questions (the difference is associated with a p value = 0.1350).[15] Furthermore, though not reported in Table 4,

[15] While Table 4 might appear to suggest that all-female rides perform less well than all-male rides, this is not actually the case. Whereas all-female groups are asked 8 % fewer questions than all-male groups, the

there is no difference across the gender groups in the number of correct answers given on the first four questions of the ride, nor is there is a difference in the rate at which the different gender groups successfully completed a Red Light Challenge. Thus, all-female groups appear to be equally adept as all-male groups at answering questions correctly in the Cash Cab. These results suggest that there were no inherent differences in ability before the ride nor were there performance differences by the end of the ride.

Pre-existing confidence is more difficult to assess. On the one hand, confidence appears to be calibrated equally well across these groups (for a discussion of confidence calibration, see Hoelzl and Rustichini 2005, Lichtenstien and Fischhoff 1977). That is, the probability of answering correctly when highly confident does not differ across groups, nor does it differ across groups when considering some or no confidence. Yet, there are significant differences across the gender groups with respect to the frequency of high, some, and no confidence responses. Table 4 indicates that all-male groups exhibit high confidence in 80 percent of their correct answers (7.15 out of 8.99), whereas all-female groups exhibit high confidence in only 71 percent of their correct answers (5.70 out of 8.03). In other words, women are less confident than men, even in their correct answers.

Of course, the more important question regarding confidence is whether contestants appear to use their in-game confidence to update their subjective probability of being able to answer the VBQ correctly. Table 4 displays that 44 percent of all-male groups risk their winnings on the VBQ gamble in our sample, compared to only 30 percent of all-female groups. One interpretation of the regression results from Table 3, however, is that men, at best, base their risky decisions on general aspects of their previous confident, good play (not all of which is relevant at the time the decision is made) and at worst fail to condition their risky decisions on any of the relevant information available to them. In sharp contrast, women appear to consider all of the information available to them, including previous poor play as well as by their most recent confident, good play, which, by design, is likely the most relevant information to consider.

These results are consistent with the idea that contestants (especially women) update their subjective probabilities based on their previous play of the game. Furthermore, the data indicate that all-female groups are more likely than all-male groups to use relevant information when updating their subjective probability of answering the VBQ correctly. Recall that (a) when presenting the VBQ to the contestants, the host of *Cash Cab* states that the VBQ is harder than previous questions, and (b) contestants are also told at the start of the game, and twice during the game, that the questions get more difficult as the money value increases (after the fourth question and again after the eighth). Given this information, contestants *should* weight their performance on harder questions (as all-female groups seem to do) more than on the easier questions rather than errantly considering their entire play of the game (as all-male groups seem to do).

Footnote 15 continued
all-female rides are also 2.3 city blocks shorter than all-male rides, which corresponds to being about ten percent shorter. Similarly, although Table 4 shows that, on average, all-female rides earn less money at the completion of their ride (i.e., before the VBQ) than all-male groups, this difference too can be attributed to the additional 0.94 questions answered correctly by all-male groups (which is typically a question worth $200) as well as to the extra frequency with which longer distance rides are presented with the opportunity to answer a Red Light Challenge.

Under this scenario, males appear to be less rational in that they do not fully consider information when making decisions, while females appear to be more rational in that they consider a wider range of the relevant information available to them.

Future research could benefit from studying the group dynamics of the decisions made on *Cash Cab* (e.g., Ertac and Gurdal 2012; Ronay and Kim 2006). Unfortunately, given that the US version of the show does not broadcast group deliberations in their entirety, such an analysis would require access to the original, unedited footage. With such access, one could consider the length of deliberations and other group dynamics when considering the VBQ offer, or even link a contestant's comments during her team's deliberations to her own personal contributions earlier in the game. Alternatively, one could use the German version of *Cash Cab*, called *Quiz Taxi*, which restricts the number of contestants on each ride to 2 or 3, and airs the entire discussion among contestants when debating whether to accept the VBQ . Indeed, using data from *Quiz Taxi*, Keldenich and Klemm (2012) show that more extensive discussions prior to the VBQ help groups arrive at the "right" decision. On a final note, access to unedited footage could determine whether contestants consider the temporal dynamics of game performance (e.g., pace of question delivery and responses; initial confidence; deliberation time; time to certainty, etc.) when considering the final gamble.

Acknowledgments We thank Sarah Cattano, Line Producer for Lion Television, for discussing production aspects of *Cash Cab*. The paper benefited from comments offered by the Editor and two anonymous referees. All remaining errors are our own.

References

Barsky, R. B., Juster, F. T., Kimball, M. S., & Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *Quarterly Journal of Economics*, *112*(2), 537–579.

Bliss, R. T., Potter, M. E., & Schwarz, C. (2012). Decision making and risk aversion in the cash cab. *Journal of Economic Behavior and Organization*, *84*(1), 163–173.

Booij, A. S., van Praag, B. M. S., & van de Kuilen, G. (2010). A parametric analysis of prospect theory's functionals for the general population. *Theory and Decision*, *68*(1–2), 115–148.

Borghans, L., Golsteyn, B. H., Heckman, J. J., & Meijers, H. (2009). Gender differences in risk aversion and ambiguity aversion. *Journal of the European Economic Association*, *7*(2–3), 649–658.

Brooks, R., Faff, R., Mulino, D., & Scheelings, R. (2009). 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*, *9*(1/2), 27–50.

Bruhin, A., Fehr-Duda, H., & Epper, T. (2010). Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica*, *78*(4), 1375–1412.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, *125*(5), 367–383.

Cohen, A., & Einav, L. (2007). Estimating risk preferences from deductible choice". *American Economic Review*, *97*(3), 745–788.

De Roos, N., & Sarafidis, Y. (2010). Decision making under risk in deal or no deal. *Journal of Applied Econometrics*, *25*(6), 987–1027.

Deck, C., Lee, J., & Reyes, J. (2008). Risk attitudes in large stake gambles: Evidence from a game show. *Applied Economics*, *40*(1), 41–52.

Donkers, B., Melenberg, B., & van Soest, A. (2001). Estimating risk attitudes using lotteries: A large sample approach. *Journal of Risk and Uncertainty*, *22*(2), 165–195.

Eckel, C. C., & Grossman, P. J. (2005). Differences in the economic decisions of men and women: Experimental evidence. In C. R. Plott & V. L. Smith (Eds.), *Handbook of experimental economics results* (Vol. 1, pp. 309–519). New York: Elsevier.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, *75*(4), 643–669.

Ertac, S., & Gurdal, M. Y. (2012). Deciding to decide: Gender, leadership and risk-taking in groups. *Journal of Economic Behavior & Organization*, *83*(1), 24–30.

Fehr-Duda, H., De Gennaro, M., & Schubert, R. (2006). Gender, financial risk, and probability weights. *Theory and Decision*, *60*(2–3), 283–313.

Finucane, M. L., Slovic, P., Mertz, C. K., Flynn, J., & Satterfield, T. A. (2000). Gender, race, and perceived risk: The 'White Male' effect. *Healthy Risk & Society*, *2*(2), 159–172.

Flynn, J., Slovic, P., & Mertz, C. K. (1994). Gender, race, and perception of environmental health risks. *Risk Analysis*, *14*(6), 1101–1108.

Fullenkamp, C., Tenorio, R., & Battalio, R. (2003). Assessing individual risk attitudes using field data from lottery games. *Review of Economics and Statistics*, *85*(1), 218–226.

Gardner, M., & Steinberg, L. (2005). Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: An experimental study. *Developmental Psychology*, *41*(4), 625–635.

Gertner, R. (1993). Game shows and economic behavior: Risk-taking on 'card sharks'. *Quarterly Journal of Economics*, *108*(2), 507–521.

Hersch, J. (1996). Smoking, seat belts and other risky consumer decisions: Differences by gender and race. *Managerial and Decision Economics*, *17*(5), 471–481.

Hersch, P. L., & McDougall, G. S. (1997). Decision making under uncertainty when the stakes are high: Evidence from a lottery game show. *Southern Economic Journal*, *64*(1), 75–84.

Hoelzl, E., & Rustichini, A. (2005). Overconfident: Do you put your money on it? *Economic Journal*, *115*(503), 305–318.

Jianakoplos, N. A., & Bernasek, A. (1998). Are women more risk averse? *Economic Inquiry*, *36*(4), 620–630.

Keldenich, K., & Klemm, M. (2012). Double or nothing!? Small groups making decisions under risk in 'quiz taxi'. *Journal of Economic Behavior and Organization*, *84*(1), 163–173.

Lichtenstien, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance*, *20*(2), 159–183.

Post, T., van den Assem, M. J., Baltussen, G., & Thaler, R. H. (2008). Deal or no deal? Decision making under risk in a large-payoff game show. *American Economic Review*, *98*(1), 38–71.

Ronay, R., & Kim, D. (2006). Gender differences in explicit and implicit risk attitudes: A socially facilitated phenomenon. *British Journal of Social Psychology*, *45*(2), 397–419.

Vinokur, A. (1971). Review and theoretical analysis of the effects of group processes upon individual and group decisions involving risk. *Psychology Bulletin*, *76*(4), 231–250.