Applied Cognitive Psychology, *Appl. Cognit. Psychol.* **27**: 542–551 (2013)
Published online 3 June 2013 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/acp.2933

Decision-making Under Uncertainty in the *Cash Cab*

MATTHEW R. KELLEY[1]* and ROBERT J. LEMKE[2]

[1]*Department of Psychology, Lake Forest College, Lake Forest, USA*
[2]*Department of Economics, Lake Forest College, Lake Forest, USA*

Summary: Television game shows have long been used to analyze risk-taking behavior. We used episodes of *Cash Cab* to investigate how a variety of pre-game and in-game factors affected contestants' decisions to accept or reject a double-or-nothing gamble offered at the end of the game. As expected, the analysis confirmed the standard influences of gender, age, and group size on the willingness to accept the gamble. More interestingly, however, the data suggested that contestants also used in-game experiences to update their subjective probability of success when considering the final gamble. Surprisingly, contestants did not appear to use correct performance (e.g., number correct, and streaks of correct) when updating, but the number and distribution of highly confident and correct responses were important when assessing the final gamble. Copyright © 2013 John Wiley & Sons, Ltd.

Televised game shows have long been attractive to researchers studying decision-making under uncertainty, in part, because they offer payouts that are significantly greater than what can typically be offered in a laboratory or classroom setting (e.g., Bliss, Potter & Schwarz, 2011; Brooks, Faff, Mulino, & Scheelings, 2009; De Roos & Sarafidis, 2010; Deck, Lee, & Reyes, 2008; Gertner, 1993; Metrick, 1995; Post, Van den Assem, Baltussen & Thaler, 2008). Much of this literature has examined the level of risk involved in a contestant's wager as a function of relatively straightforward variables such as the amount of money they had won and the objective mathematical probabilities of their future success. This approach has worked well for shows such as *Card Sharks* (e.g., Gertner, 1993) and *Deal or No Deal* (e.g., De Roos and Sarafidis, 2010) because it is not difficult for contestants (or researchers) to determine the objective probability of success simply by determining the number of high value cards or suitcases remaining in the game. Of central interest in this paper, however, is the question of how contestants make decisions in the absence of such objective probabilities of success. That is, what factors do contestants consider when they must rely on a subjective probability of their future success?

To this end, we considered a televised trivia game show (*Cash Cab*) that has three important features: (i) as it is a trivia game show, the objective probability of success is unknown; (ii) a contestant makes several observable decisions while playing the main part of the game, the outcomes of which might indicate whether contestants update their subjective probability of success on the basis of what they have learned during the game; and (iii) the game ends by presenting the contestant with a risky offer that he or she can accept or reject. The design of *Cash Cab* allows us to examine the influence of a variety of factors on a contestant's decision whether to risk his or her winnings on the final gamble, including pre-game (e.g., gender, race, and age) and in-game factors (e.g., winnings, accuracy, confidence, and streaks). Accordingly, in the absence of an objective probability of success, we should be able to assess the extent to which contestants rely on factors that were determined before even setting foot in the *Cash Cab*, as well as whether they use aspects of their game play to update their subjective probabilities of success. Whereas the influence of pre-game factors has been well documented in a variety of uncertainty paradigms, our novel contribution concerns the latter issue—does the in-game experience influence the updating of subjective probabilities and the subsequent risky decision? Because the predictions offered in the paper are closely tied to the features of the game show, we will begin with a detailed description of *Cash Cab*.

TAKING A RIDE IN THE *CASH CAB*

Cash Cab, hosted by Ben Bailey in Manhattan (NY), aired on the Discovery Channel in the USA from 2005 to 2012. According to the Discovery Channel website (2012), 'unassuming people enter the "Cash Cab" as simple passengers taking a normal taxi ride, only to be shocked when they discover that they're instant contestants.' That is, the show purports that it does not prescreen or preselect its contestants, but instead, it surprises potential passengers with the opportunity to win money by answering a series of trivia questions during their cab ride. Upon entering the cab, the passengers, from as few as one to as many as four, are given the opportunity to leave the cab and not play (which some do) or to start the game. The game ends either when (i) the contestants accumulate three strikes by answering three questions incorrectly, in which case no money is won and the contestants are asked to leave the cab immediately, or (ii) the cab reaches the contestants' destination, in which case the contestants are entitled to all the money they have accumulated. The first set of four questions of each game (or ride) is worth an initial value of $25, $50, or $100, depending on the production season and whether the passengers have been randomly selected for a Cash Cab Double Ride. The value of the next set of four questions (Q5–Q8) is double the initial value, whereas all remaining questions (Q9 and above) are worth quadruple the initial value.

In addition to these basic premises, there are three notable aspects to the game. First, contestants can ask for help twice in the game either by using their *mobile shout-out* (call a friend) or their *street shout-out* (invite a passerby over to the cab). Second, the first time the cab stops at a red light

*Correspondence to: Matthew Kelley, Department of Psychology, Lake Forest College, Lake Forest, IL 60035, USA.
E-mail: kelley@lakeforest.edu

after the contestants have accumulated at least $250 in winnings, the contestants are given a *Red Light Challenge* (*RLC*) in which they have 30 seconds to answer all parts of a multiple-response question (e.g., name all US states that border the Pacific Ocean). Contestants are not allowed to use shout-outs on the RLC, and there is no penalty for a failed RLC attempt (i.e., no strike). The RLC is worth $250 if answered correctly for rides with an initial question value of $25 or $50 and is worth $500 for rides with an initial question value of $100.

Third, contestants who successfully complete their ride (i.e., reach their destination before receiving three strikes) are presented with a risky proposition: they can take their accumulated winnings and end the game, or they can risk all of their winnings on a double-or-nothing gamble. To win the gamble, contestants must correctly answer a *video bonus question* (*VBQ*) that is, according to the host, more difficult than the previously asked questions and without the aid of a shout-out. If they answer the VBQ correctly, their winnings are doubled; if they answer incorrectly, they lose all previous winnings and exit the cab at their destination having won nothing (except a free cab ride).

The primary analysis of this paper concerns this third aspect of the game, which allows us to examine decision-making under uncertainty. That is, contestants who successfully reached their destination had, roughly 15 to 30 minutes earlier, no idea that they would be on a game show and be offered a gamble that asked them to choose between (i) a safe option in which they are guaranteed some money and (ii) a risky double-or-nothing option that depends on answering a single question correctly. Being presented with this gamble, therefore, allows us to investigate risk. In terms of this analysis, one particularly convenient feature of *Cash Cab* is that the variability in the starting dollar values ($25, $50, and $100) of the rides ensures that the winnings are not necessarily closely tied to the contestant's performance. For example, completing one's ride with $1200 could be achieved in any number of ways, such as by answering all 17 questions correctly without the use of shout-outs on a $25 ride or by answering seven of nine questions correctly while using both shout-outs on a $100 ride. Hence, we might be able to tease apart underlying differences in risk-taking behavior even for people with the same accumulated winnings as a function of other factors involved in their ride.

CONSIDERING THE GAMBLE

In the absence of a known objective probability of success, what factors do *Cash Cab* contestants consider when they decide whether to accept or reject the final gamble? Before even entering the cab, contestants are equipped with a set of beliefs about (and experience with) risky decisions, but, unfortunately, the structure of the show does not allow us to assess these preexisting risk attitudes. Without this information, we must rely on indirect indicators of preexisting risk attitudes, such as the contestant's gender, race/ethnicity, and age. Past research has shown that men are more likely to take risks than women (for a meta-analysis, see Byrnes, Miller & Schafer, 1999), whites tend to perceive less risk than nonwhites (e.g., Finucane, Slovic, Mertz, Flynn, & Satterfield, 2000; Flynn, Slovic & Mertz, 1994), and the willingness to take risks decreases as people grow older (e.g., Barsky, Juster, Kimball & Shapiro, 1997; Donkers, Melenberg & Van Soest, 2001). In addition, the presence of other contestants and the size of the group influence risky decisions, as individuals tend to take more risks when in groups than when alone (e.g., Gardener & Steinberg, 2005; Keldenich and Klemm, 2011; Vinokur, 1971). Given the prevalence of these findings in the literature, we offer the general prediction that these *pre-game indicators* will be important when considering risk-taking behavior in the *Cash Cab*. More specifically, we offer four subhypotheses: (1a) men will be more likely to accept the gamble than women; (1b) whites will be more likely to accept the gamble than nonwhites; (1c) the probability of accepting the gamble will decrease with age; and (1d) the probability of accepting the gamble will increase with group size.

Standard risk theory contends that people accept less risk when more is at stake. Simply put, people are generally risk averse, especially when the monetary amount at risk becomes large as is commonly the case with game shows. In line with previous game show research, therefore, we expect that the amount of winnings at risk (the *monetary factor*) will be an important consideration for contestants. Specifically, our second hypothesis is (2) contestants will be less likely to accept the gamble the more money they have won at the completion of the ride.

As mentioned earlier, the primary goal of the present analysis is to determine whether contestants use aspects of their game play to determine and update their subjective probabilities of success. Specifically, subjective probabilities may be a function of a variety of performance factors. Many researchers have shown that risk-seeking behavior increases following prior gains, which are akin to correct answers in the *Cash Cab* (e.g., Kahneman & Tversky, 1979; Thaler & Johnson, 1990). Moreover, contestants might be sensitive to their streaks in accuracy or errors. Having long streaks of successes during the game or ending the game with a streak of correct answers might increase the likelihood the contestants accept the final gamble, much like belief in the *hot hand* in sports can influence shot frequency, distribution, and selection (for a review, see Bar-Eli, Avugos & Raab, 2006).[1] Also, the contestant's performance on the last few questions of the ride may have a greater influence on the decision to gamble than things that happened early in the ride if they are given more weight in the updating process; these events are likely to be more memorable (recency bias), and they may effectively serve as an anchor when deciding to gamble (e.g., Ahlawat, 1999; Hogarth & Einhorn, 1992). Later questions might also deserve more weight in the *Cash Cab* setting, as contestants are told that question difficulty increases throughout the game and that the VBQ is harder

[1] The Bar-Eli, Avugos, and Raab (2006) review showed that most studies have failed to provide evidence that the *hot hand* phenomenon exists. Similarly, we failed to find any evidence of a 'hot hand' for trivia performance while in the *Cash Cab*—the streak data failed to differ significantly from streaks expected purely by chance. Despite these failures to produce the phenomenon, people still tend to believe that it exists, and this belief can influence their subsequent behavior.

still—indeed, the data bore this out as performance dropped from 85% to 74% to 73% across blocks Q1–Q4, Q5–Q8, and Q9+, respectively. We expect several *in-game performance factors* to increase the likelihood of accepting the final gamble. Specifically, we hypothesize that the likelihood of accepting the final gamble will increase: (3a) as the overall number of correct answers given increases; (3b) as the length of the longest streak of correct answers increases; (3c) as the length of the ending correct streak increases; and (3d) when contestants answer the final question correctly.

Although correct answers are necessary to avoid strikes and to keep playing the game, when it comes to updating beliefs when confronted with the gamble, the contestant's confidence in previous answers might also prove important in making the final decision. For instance, imagine two separate rides that both ended with eight correct answers out of eight questions where contestant A was highly confident in each response and contestant B was never highly confident, but instead made some lucky guesses and received help via shout-outs. One might expect that contestant A would be more likely to accept the final gamble than contestant B despite the fact that their accuracy and winnings were identical. Indeed, confidence and the proper calibration of confidence have proven important when making decisions (e.g., Griffin & Tversky, 1992; Lichtenstein & Fischhoff, 1977; Nosic & Weber, 2010). Therefore, we expect several *in-game confidence factors* to increase the likelihood of accepting the final gamble. Specifically, we hypothesize that the likelihood of accepting the final gamble will increase: (4a) as the overall percentage of highly confident and correct (HC) answers increases; (4b) as the length of the longest HC streak increases; (4c) as the length of the ending HC streak increases; and (4d) when contestants answer the final question correctly with high confidence.

METHOD

Sample

Data were collected from the 196 episodes of *Cash Cab* that aired during the 26 weeks of 15 May 2011 through 12 November 2011. Eight episodes from the first season of *Cash Cab* were omitted from the analysis because they were played slightly differently than the rest. In particular, the host offered more help to the contestants in these early shows by reminding them of the rules of the game and encouraging the use of shout-outs.

Data collection

Two types of data—objective (e.g., response accuracy) and subjective (e.g., response confidence)—were collected by having both authors watch each show in its entirety. 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 through discussion, and the observation was flagged to indicate a lack of initial inter-rater agreement.

In advance of the first trivia question, the following objective measures were recorded: the sex of each contestant, the number of contestants, and the initial value of the questions. All except two shows included three 'rides' of separate sets of contestants. Each ride consisted of one, two, three, or four contestants. Although babies were not included in the number of contestants on a particular ride, children who could participate were included; indeed, children often contributed correct answers on the rides. During the trivia portion of the game itself, objective data were recorded on each question as to whether the contestants provided an accurate response and whether a shout-out was used. If the contestants were offered a RLC, we recorded when in the game it occurred, its value, and whether the RLC was answered correctly.

Further, we recorded a variety of subjective measures throughout the ride. At the beginning, we categorized each contestant in terms of race (Asian, Black, Hispanic, Native American, Pacific Islander, or White) and age (younger than 18, 18–30, 31–50, 51–65, and older than 65 years). We also recorded which, if any, contestants indicated prior knowledge of the show. Immediately after the contestant gave his or her final answer to a question, we rated the contestant's confidence in his or her answer by assigning it to one of three categories: highly confident, somewhat confident, and not at all confident. Confidence was assessed by considering contestants' statements, tone, and body language immediately prior to and as they answered the question and not by whether the contestants actually answered a question correctly or not.[2] When contestants made statements such as 'I have no idea' or 'I don't even have a good guess', they were rated as having no confidence; these statements were often paired with shrugs of shoulders, shaking of heads, eye rolls, and long, slow exhalation of breath. In contrast, contestants who offered statements such as 'I know this', 'I've got this one', or 'I'm totally confident' were rated as having high confidence; these reports were often paired with smiles, sitting up straight in their seats, fist pumps, claps, and high fives. Finally, ratings of some confidence stemmed from declarations such as 'Hmm, I'm pretty sure' or 'I'm not certain but…' along with furrowed brows, gritted teeth, held breath, crossed fingers, tightened shoulders, and sitting forward in seats as they waited to hear the answer.[3]

Contestants who failed to reach their destination with a positive money balance because they received three strikes (three wrong answers) were recorded as having left the cab with $0. In contrast, for those contestants who successfully reached their destination with a positive money balance, their balance was recorded as was their decision of whether to gamble their winnings on the VBQ, whether they answered the VBQ correctly, and their total winnings upon leaving the cab.

[2] The authors' independent assessment of confidence matched on over 93% of all questions asked of the contestants. In the final dataset of all rides, 96% of answers given with high confidence were correct, 59% of answers given with some confidence were correct, and 5% of answers given with no confidence were correct. For all completed rides, the percentage of answers given correctly was 97, 69, and 9, respectively.

[3] The host's feedback on whether the contestant's answer was correct did not occur immediately after the contestant's response. The typical delay was between 1 and 5 seconds, which was ample time to pause the recording before learning the accuracy of the response.

Table 1. Descriptive statistics

	Risked winnings on the VBQ				Did not risk winnings on the VBQ			
	($N = 159$)				($N = 220$)			
	Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
Percentage of riders who are female	41.1	33.7	0	100	48.0	36.3	0	100
Percentage of riders who are white	88.0	25.5	0	100	85.5	30.6	0	100
Average age of riders	33.9	8.5	26	57	37.3	10.0	26	70
Number of riders	2.6	0.9	1	4	2.3	0.8	1	4
Number of questions asked	9.8	2.2	5	16	10.1	2.5	6	19
Money winnings before the video bonus challenge	790	413	225	3100	904	467	175	2900
Number of correct answers given	8.4	2.3	3	15	8.7	2.6	4	19
Number of shout-outs used	0.76	0.72	0	2	0.89	0.73	0	2
Number of answers given with high confidence	7.21	2.42	1	13	6.80	2.69	1	16
Number of answers given with some confidence	1.42	1.12	0	6	1.96	1.37	0	6
Number of answers given with no confidence	1.12	0.90	0	4	1.36	0.99	0	4
Longest streak of correct answers	6.33	2.52	2	15	6.47	2.75	2	19
Streak of correct answers to end the game	3.33	3.29	0	15	3.21	3.27	0	19
Answered the last question correctly	0.862	0.346	0	1	0.827	0.379	0	1

Source: Authors' calculations from self-collected data from *Cash Cab* episodes that aired on the Discovery Channel from 15 May 2011 through 12 November 2011.

RESULTS

Descriptive statistics and selection checks

The 196 episodes of *Cash Cab* yielded data on 586 rides; 379 (64.7%) of these rides were *completed rides* in which the contestants were given the option of taking the final gamble. Of these, 159 (42%) accepted the gamble to risk their winnings, whereas 220 (58%) declined the gamble. Table 1 displays the summary statistics for the key variables used in the analysis, separately for (i) all completed rides that risked their winnings and (ii) all completed rides that did not risk their winnings.

Given that the primary analyses in this paper concern whether contestants used their in-game performance and confidence to update their subjective probability of success when considering the gamble, we must first determine whether there were initial differences between the eventual gamblers and nongamblers in terms of ability (performance), confidence, and prior knowledge of the game. Overall, a series of independent-samples *t*-tests and chi-square tests for independence suggest that these two groups did not differ on these characteristics. To assess ability, we examined (i) performance on the RLC—a riskless question that could not be counted as a strike—and (ii) the number of correct responses across the first block of four questions. In both cases, we found no significant differences between the groups ($ps = 0.731$ and 0.670, respectively). Moreover, we found no significant differences between these groups across the first four questions in terms of the number of highly confident responses ($p = 0.279$). Finally, eventual gamblers and nongamblers did not differ in their self-reported prior knowledge of the show ($p = 0.951$). Moreover, there was also no difference in the rate of use of shout-outs in the first four questions between gamblers and nongamblers ($p = 0.677$), which might be indicative of equal ability across the groups or that the groups are not playing the game differently early on. Hence, the data suggest that gamblers do not start the game any more accurate or confident than do nongamblers, nor do they have more prior knowledge of the game than do nongamblers. Of course, we are predicting that, by the end of the game, accuracy and confidence will ultimately influence the decision to gamble but that this influence is based on what contestants learn during their game play, as opposed to simply beginning the game inherently more intelligent and/or more confident.

Regression results: pre-game indicators

Although the focus of this paper does not concern the effect of pre-game factors on risk behavior, it is important to show that the *Cash Cab* data support previous findings. To test our four hypotheses regarding pre-game indicators (Hypotheses 1A–1D), we estimated the decision of whether to accept the gamble to be a logistic function of pre-game variables: percentage of contestants on each ride who are female, percentage of contestants on each ride who are white, the average age of all contestants on each ride, the number of contestants on each ride, and the contestant's winnings that would be at risk if the gamble is accepted. Table 2 reports the estimated marginal effects and *p*-values for each of these variables.[4]

The first four rows of Table 2 show that risk behavior in the *Cash Cab* regarding pre-game factors largely mirrors what researchers have found in other settings. In particular, the more male contestants on a ride, the more likely the gamble is to be accepted ($p = 0.013$); the older the contestants are on average, the less likely the gamble is to be accepted ($p = 0.004$); and the gamble is more likely to be accepted as the number of contestants increases ($p < 0.001$). The only pre-game factor that is not statistically related to risk behavior is race. Although white contestants are more likely to

[4] To save space, the regression results are not reported, but they are available from the corresponding author upon request. Moreover, the estimated marginal effects from several other models that include both pre-game and in-game factors are reported in the Appendix tables. In each case, the implied *p*-value for each pre-game factor is fairly close to the *p*-value reported in Table 2.

Table 2. Hypothesis results

	Estimated marginal effect (%)	p-value	Confirmed?
Hypothesis 1: pre-game indicators			
A. Men riskier than women	−19.64	0.013	◀
B. Whites riskier than nonwhite	10.11	0.295	No
C. Risk decreases as age increases (per year)	−0.86	0.004	◀
D. Risk increases with number of riders	12.07	<0.001	◀
Hypothesis 2: amount at risk (winnings per $100)	2.24	0.001	◀
Hypothesis 3: in-game performance			
A. Risk increases with total number correct	0.05	0.969	No
B. Risk increases with longest streak correct	0.51	0.648	No
C. Risk increases with last streak correct	1.17	0.182	No
D. Risk increases if last question correct	11.63	0.106	No
E. Joint: All		0.410	No
Hypothesis 4: in-game confidence			
A. Risk increases with total number of HC answers	2.74	0.024	◀
B. Risk increases with longest HC answer streak	3.32	0.011	◀
C. Risk increases with last HC answer streak	4.25	0.005	◀
D. Risk increases if last question was HC	12.48	0.022	◀
E. Joint: All		0.002	◀

Notes: The p-values testing Hypotheses 1 and 2 are from a multiple variable logistic regression; the results of which are available from the authors upon request. The p-values testing Hypotheses 3 and 4 are from the multiple variable logistic regressions; the results of which are provided to the reader in Appendix Tables A1 and A2. HC, highly confident and correct.

accept the gamble than are nonwhites, the difference is not statistically significant ($p = 0.295$).

Hypothesis 2 predicted that contestants would be less likely to accept the gamble the more money would be at risk. This hypothesis was also confirmed at a p-value of 0.001, as reported in the fifth row of Table 2. Thus, overall, our *Cash Cab* data supported our hypotheses regarding pre-game indicators (1A, 1C, and 1D) and the monetary factor (2).

Regression results: in-game factors

The primary contribution of this paper comes from an analysis of how contestants allow in-game factors to affect risk behavior, potentially by influencing subjective probabilities of success. Hypothesis 3 proposed that the number of correct answers given by a contestant might positively influence the decision to accept the gamble. Specifically, we consider four possible measures of in-game success: (3A) total number of correct answers; (3B) longest streak of correct answers; (3C) the number of correct answers to end the game; and (3D) answering one's final question correctly.[5]

The average ride on *Cash Cab* only asked 10 questions (getting about 8.5 correct on average). Thus, these four measures are highly correlated. In order to test the third set of hypotheses, five logistic regressions were carried out—one with each measure of accuracy plus one regression that included all four measures. The estimated marginal effects for all five models are reported in Appendix A1. Although all coefficient estimates are positive, none of the four measures of accuracy are statistically significant predictors of accepting the gamble. The p-value from each measure is reported in Table 2 (rows 3A–3D); none of which are below the 10% significance level. Further, the fifth logistic regression model reported in Appendix A1 includes all four measures in order to test if response accuracy, measured in all of its forms, can help explain the decision to accept the gamble. A joint test that all accuracy measures have no effect was carried out, and the p-value of this joint test is 0.410 (Table 2; row 3E). Thus, taken individually or collectively, measures of accuracy do not help predict behavior. Counter-intuitively, there is a striking lack of support for Hypothesis 3—contestants on *Cash Cab* do not appear to rely on response accuracy when deciding whether to accept the gamble.

As a further robust check, Appendices B1 and C1 report marginal effects from estimating the same model as described earlier but under probit regression (B1) and under a linear probability model with corrected standard errors (C1). The results are strikingly similar across all models. Of the 12 models that include a single measure of accuracy, accuracy was marginally statistically significant only once with a $p = 0.091$ (last question correct; linear probability model).

Whereas Hypothesis 3 considered only response accuracy, Hypothesis 4 maintained that confidence would influence the decision to gamble. Specifically, we considered four possible measures of in-game success augmented with confidence: (4A) total number of highly confident correct answers; (4B) longest streak of highly confident correct answers; (4C) the number of highly confident, correct answers to end the game; (4D) and answering one's final question correctly with high confidence. In order to test the fourth set of hypotheses, five logistic regressions were carried out—one with each measure of confidence-augmented success plus one regression that includes all four measures. The marginal effects for all five models are reported in Appendix A2.

The p-values are reported in Table 2 (rows 4A–4D). Unlike accuracy measures, all four measures of confidence were statistically significant at the 5% level. Specifically, contestants were more likely to accept the gamble as the number of highly confident, correct answers given increased ($p = 0.024$); as the longest streak of highly confident, correct

[5] We model contestant behavior on a per-question basis, because the average ride is asked only 10 questions. All of the results are qualitatively unchanged if we model behavior in terms of percentages (e.g., the percent of questions answered correctly), with each question representing about 10% of a contestant's questions. Regression results using percentages are available upon request.

answers increased ($p = 0.011$); as the length of the ending streak of highly confident, correct answers increased ($p = 0.005$); and when the final question was answered correctly with high confidence ($p = 0.022$).

To further test Hypothesis 4, the fifth logistic regression model reported in Appendix A2 included all four measures. A joint test that all confidence measures simultaneously had no effect was then carried out. Unlike in Models (1)–(4) of Table A2 (p-values reported in rows 4A–4D of Table 2), none of the marginal effects were statistically significant in the full model. But this is not unexpected, as the four confidence measures are highly correlated, and therefore, all of the standard errors increased substantially, nullifying the statistical significance of any one variable. Still, the p-value of the joint test that none of the confidence measures is correlated with the decision to accept the gamble was rejected with a p-value of 0.002 (reported in row 4E in Table 2). Although the degree to which any particular confidence variable affects the decision is unclear because of bloated standard errors, the conclusion remains that confidently providing correct answers during the ride is positively related to accepting the gamble. Thus, the support for Hypothesis 4 was overwhelming: confidence in answering questions correctly during the game had an important influence on decision-making.

As a further robust check, Appendices B2 and C2 report the marginal effects from estimating the same model as described earlier, but under probit regression (B2) and under a linear probability model with corrected standard errors (C2). The results are strikingly similar across all models and estimation techniques.

DISCUSSION

The present study was designed to examine decision-making under uncertainty in the context of the trivia game show *Cash Cab*. Using data from 379 completed rides, we explored the influences of *pre-game* (e.g., gender, race, age, and group size) and *in-game* (e.g., winnings, accuracy, and confidence) factors on the decision to accept or reject a risky double-or-nothing proposition. Consistent with past research from a variety of paradigms (e.g., Barsky et al., 1997; Byrnes et al., 1999; Flynn et al., 1994; Vinokur, 1971), three of the four *pre-game indicators* proved to be important influences on the decision to gamble; only the effect of race failed to reach statistical significance. Men, groups, and younger adults were more likely to gamble than women, individuals, and older adults, respectively. Although these results were completely expected and not particularly novel, it is reassuring to see that contestants in the *Cash Cab* show these standard effects.

After controlling for these pre-game effects, we were able to consider the question of central interest in this paper: do contestants use aspects of their game play to update their subjective probabilities of success when considering the decision to gamble? Not surprisingly, and congruent with past research (e.g., Gertner, 1993; Metrick, 1995), contestants were less willing to accept the gamble as the amount of winnings at risk increased. More interestingly, however, we found that in-game performance did not influence the decision to gamble. Whether measured by the total number of correct answers, the longest streak correct, the ending streak correct, or whether the final answer was correct, contestants did not appear to consider accurate performance when updating their subjective beliefs about success.

Whereas these results initially seemed counterintuitive, they made more sense when we considered the joint influence of being HC. The probability of accepting the gamble increased significantly as the number of HC answers increased, as well as when contestants had long streaks of HC answers, ended the game with a long streak of HC answers, and even when just the final question was answered correctly with high confidence. Controlling for all four of these factors simultaneously gives just a slightly different picture. Because of the small sample size and high degree of colinearity across the confidence measures, none of the measures were individually significant predictors of behavior. Yet the joint test revealed that confidence, measured in one form or another, was highly significant. Given that eventual gamblers and nongamblers were indistinguishable on measures of initial ability and confidence, we believe these results reflect that contestants pay particular attention to their in-game confidence—and perhaps how well their confidence is calibrated—and use this information to update their subjective probability of success when deciding whether to gamble. In other words, contestants learn during their game play and factor this information into their final decision.

Although the present research has identified a variety of factors that affect subjective probability and predict a subsequent gamble, future research with *Cash Cab* might identify further variables of importance. For instance, with access to the unedited versions of the rides, one could examine whether the temporal dynamics of game performance (e.g., pace of question delivery and responses, initial confidence, deliberation time, and time to certainty) influence subjective probabilities. Imagine two rides of 13 questions each where Ride 1 was completed over 16 blocks in 5 minutes and Ride 2 was completed over 42 blocks in 20 minutes. Even with identical performance, confidence, and streaks, Ride 1 is likely to feel different to the contestants—crisper pace, less deliberation, and more initial certainty—than Ride 2. One might not be surprised, then, if the contestants in Ride 1 were more likely to take the gamble once these temporal dynamics are considered. In addition, with unedited footage, one could consider the length of deliberations and other group dynamics when considering the VBQ offer or even tie a contestant's comments during his or her team's deliberations to his or her own personal contributions earlier in the game (e.g., if one member of the team did not offer any correct answers during the course of the ride, he or she might be less wanting to accept the gamble and therefore try to convince his or her teammates to reject the gamble).[6]

[6] The German version of *Cash Cab*, called *Quiz Taxi*, is slightly different than the American version. In particular, it restricts the number of contestants on each ride to two or three, and the show airs the entire discussion among contestants when debating whether to accept the VBQ. Keldenich and Klemm (2011) use this extra footage to investigate the role group dynamics play in the decision.

Finally, we must take caution when attempting to generalize these results beyond the present study. Although we were able to measure a myriad of factors here and replicated most standard results, we had no access to a host of pre-game factors that might also influence decision-making under uncertainty, such as socioeconomic status, preexisting risk preference, inherent ability, and confidence. Future research will need to better measure these factors *a priori* and then could directly manipulate in-game factors (e.g., performance, confidence) experimentally to determine whether these factors continue to influence the decision to gamble. That said, the present study represents an important first step towards understanding these factors.

REFERENCES

Ahlawat, S. S. (1999). Order effects and memory for evidence in individual versus group decision making in auditing. *Journal of Behavioral Decision Making*, *12*, 71–88.

Bar-Eli, M., Avugos, S., & Raab, M. (2006). Twenty years of “hot hand” research: Review and critique. *Psychology of Sport & Exercise*, *7*, 525–553.

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*, 537–579.

Bliss, R. T, Potter, M. E., & Schwarz, C. (2011). Decision making and risk aversion in the Cash Cab. *Social Science Research Network* working paper, Abstract number 1818145.

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*, 27–50.

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

Cash Cab. (2012). Retrieved December 17, 2012, from http://dsc.discovery.com/tv/cash-cab.

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

Deck, C., & Lee, J., & Reyes, J. (2008). Risk attitudes in large stake gambles: Evidence from a game show. *Applied Economics*, *40*, 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*, 165–195.

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*, 159–172.

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

Gardener, 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*, 625–635.

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

Griffin, D. W., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, *24*, 411–435.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: the belief adjustment model. *Cognitive Psychology*, *24*, 1–55.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Keldenich, K., & Klemm, M. (2011). Double or nothing!? Small groups making decisions under risk in “Quiz Show.” *Ruhr Economic Papers*. 278.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? The calibration of probability judgments. *Organizational Behavior and Human Performance*, *20*, 159–183.

Metrick, A. (1995). A natural experiment in *Jeopardy*. *American Economic Review*, *85*, 240–253.

Nosic, A., & Weber, M. (2010). How riskily do I invest? The role of risk attitudes, risk perceptions, and overconfidence. *Decision Analysis*, *7*, 282–301.

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

Thaler, R. H., & Johnson, E. J. (1990). Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science*, *36*, 643–660.

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

APPENDIX

APPENDIX A1. Marginal logit effects: Explaining risk with correct answers

	(1)	(2)	(3)	(4)	(5)
Female rides (%)	−0.0020**	−0.0019**	−0.0020**	−0.0020**	−0.0020**
	0.0008	0.0008	0.0008	0.0008	0.0008
White riders (%)	0.0010	0.0010	0.0010	0.0010	0.0010
	0.0010	0.0010	0.0010	0.0010	0.0010
Average age of riders	−0.0086***	−0.0085***	−0.0085***	−0.0087***	−0.0086***
	0.0030	0.0030	0.0030	0.0030	0.0030
Number of riders	0.1208***	0.1215***	0.1221***	0.1231***	0.1236***
	0.0327	0.0326	0.0326	0.0327	0.0328
Winnings before VBQ divided by $100	−0.0226***	−0.0239***	−0.0259***	−0.0242***	−0.0254***
	0.0080	0.0080	0.0080	0.0070	0.0080
Number of correct answers	0.0005				−0.0155
	0.0127				0.0194
Longest streak of correct answers		0.0051			0.0172
		0.0113			0.0190
Streak of correct answers to end the game			0.0117		0.0029
			0.0088		0.0109
Last answer was correct				0.1163	0.1286
				0.0681	0.0824
Pseudo-R^2	0.0765	0.0769	0.0799	0.0817	0.0844

Notes. The dependent variable equals 1 or 0 if the contestant accepted or rejected the gamble of the video bonus question, respectively. Standard errors are reported below coefficient estimates. There were 379 observations.
***significant at the 1% level.
**significant at the 5% level.
*significant at the 10% level.

APPENDIX A2. Marginal logit effects: Explaining risk with highly confident and correct answers

	(1)	(2)	(3)	(4)	(5)
Female rides (%)	−0.0018**	−0.0018**	−0.0019**	−0.0019**	−0.0018**
	0.0008	0.0008	0.0008	0.0008	0.0008
White riders (%)	0.0008	0.0009	0.0010	0.0010	0.0009
	0.0010	0.0010	0.0010	0.0010	0.0010
Average age of riders	−0.0088***	−0.0090***	−0.0092***	−0.0088***	−0.0093***
	0.0030	0.0030	0.0030	0.0030	0.0030
Number of riders	0.1186***	0.1160***	0.1241***	0.1195***	0.1188***
	0.0328	0.0329	0.0329	0.0328	0.0332
Winnings before VBQ divided by $100	−0.0307***	−0.0283***	−0.0281***	−0.0256***	−0.0327***
	0.0080	0.0080	0.0070	0.0070	0.0080
Number of highly confident and correct answers	0.0274**				0.0049
	0.0121				0.0192
Longest streak of highly confident and correct answers		0.0332**			0.0216
		0.0131			0.0210
Streak of highly confident and correct answers to end game			0.0425***		0.0240
			0.0152		0.0196
Last answer was highly confident and correct				0.1248**	0.0614
				0.0538	0.0702
Pseudo-R^2	0.0866	0.0891	0.0923	0.0868	0.1005

Notes. The dependent variable equals 1 or 0 if the contestant accepted or rejected the gamble of the video bonus question, respectively. Standard errors are reported beneath coefficient estimates. There were 379 observations.
***Significant at the 1% level.
**Significant at the 5% level.
*Significant at the 10% level.

APPENDIX B1. Marginal probit effects: Explaining risk with correct answers

	(1)	(2)	(3)	(4)	(5)
Female rides (%)	−0.0019**	−0.0019**	−0.0019**	−0.0019**	−0.0020**
	0.0008	0.0008	0.0008	0.0008	0.0008
White riders (%)	0.0010	0.0010	0.0010	0.0010	0.0010
	0.0010	0.0010	0.0010	0.0010	0.0010
Average age of riders	−0.0086***	−0.0085***	−0.0084***	−0.0086***	−0.0085***
	0.0029	0.0029	0.0029	0.0029	0.0029
Number of riders	0.1155***	0.1160***	0.1160***	0.1175***	0.1175***
	0.0314	0.0314	0.0313	0.0313	0.0314
Winnings before VBQ divided by $100	−0.0203***	−0.0215***	−0.0231***	−0.0221***	−0.0227***
	0.0070	0.0070	0.0070	0.0060	0.0070
Number correct answers	−0.0006				−0.0163
	0.0124				0.0192
Longest streak of correct answers		0.0040			0.0172
		0.0110			0.0188
Streak of correct answers to end the game			0.0102		0.0016
			0.0085		0.0107
Last answer was correct				0.1137	0.1309
				0.0687	0.0834
Pseudo-R^2	0.0754	0.0756	0.0781	0.0803	0.0828

Notes. The dependent variable equals 1 or 0 if the contestant accepted or rejected the gamble of the video bonus question, respectively. Standard errors are reported beneath coefficient estimates. There were 379 observations.
***Significant at the 1% level.
**Significant at the 5% level.
*Significant at the 10% level.

APPENDIX B2. Marginal probit effects: Explaining risk with highly confident and correct answers

	(1)	(2)	(3)	(4)	(5)
Female rides (%)	−0.0017**	−0.0018**	−0.0019**	−0.0018**	−0.0018**
	0.0008	0.0008	0.0008	0.0008	0.0008
White riders (%)	0.0009	0.0009	0.0010	0.0011	0.0010
	0.0010	0.0010	0.0010	0.0010	0.0010
Average age of riders	−0.0087***	−0.0089***	−0.0089***	−0.0087***	−0.0091***
	0.0029	0.0029	0.0029	0.0029	0.0029
Number of riders	0.1122***	0.1096***	0.1165***	0.1127***	0.1099***
	0.0314	0.0315	0.0314	0.0313	0.0316
Winnings before VBQ divided by $100	−0.0277***	−0.0260***	−0.0254***	−0.0236***	−0.0296***
	0.0070	0.0070	0.0070	0.0060	0.0070
Number highly confident and correct answers	0.0265**				0.0041
	0.0118				0.0187
Longest streak of highly confident and correct answers		0.0326**			0.0219
		0.0128			0.0206
Streak of highly confident and correct answers to end game			0.0398***		0.0206
			0.0145		0.0187
Last answer was highly confident and correct				0.1226**	0.0667
				0.0532	0.0682
Pseudo-R^2	0.0851	0.0879	0.0902	0.0855	0.0985

Notes. The dependent variable equals 1 or 0 if the contestant accepted or rejected the gamble of the video bonus question, respectively. Standard errors are reported beneath coefficient estimates. There were 379 observations.
***Significant at the 1% level.
**Significant at the 5% level.
*Significant at the 10% level.

APPENDIX C1. LPM effects: Explaining risk with correct answers

	(1)	(2)	(3)	(4)	(5)
Female rides (%)	-0.0018^{**}	-0.0018^{**}	-0.0018^{**}	-0.0019^{***}	-0.0019^{***}
	0.0007	0.0007	0.0007	0.0007	0.0007
White riders (%)	0.0010	0.0009	0.0010	0.0009	0.0010
	0.0008	0.0008	0.0008	0.0008	0.0009
Average age of riders	-0.0077^{***}	-0.0076^{***}	-0.0075^{***}	-0.0077^{***}	-0.0076^{***}
	0.0025	0.0025	0.0025	0.0025	0.0025
Number of riders	0.1086^{***}	0.1092^{***}	0.1094^{***}	0.1105^{***}	0.1105^{***}
	0.0301	0.0301	0.0302	0.0301	0.0304
Winnings before VBQ divided by \$100	-0.0183^{***}	-0.0194^{***}	-0.0211^{***}	-0.0120^{***}	-0.0204^{***}
	0.0068	0.0066	0.0068	0.0061	0.0072
Number of correct answers	−0.0013				−0.0157
	0.0111				0.0173
Longest streak of correct answers		0.0033			0.0157
		0.0098			0.0169
Streak of correct answers to end the game			0.0098		0.0019
			0.0081		0.0103
Last answer was correct				0.1077^{*}	0.1232
				0.0635	0.0805
Constant	0.5956^{***}	0.5707^{***}	0.5684^{***}	0.5087^{***}	0.5188^{***}
	0.1618	0.1551	0.1471	0.1512	0.1692
R^2	0.0975	0.0977	0.1011	0.1036	0.1068

Notes. The dependent variable equals 1 or 0 if the contestant accepted or rejected the gamble of the video bonus question, respectively. Robust standard errors are reported beneath coefficient estimates. There were 379 observations.
***Significant at the 1% level.
**Significant at the 5% level.
*Significant at the 10% level.
LPM, Linear Probability Estimation.

APPENDIX C2. LPM effects: Explaining risk with highly confident and correct answers

	(1)	(2)	(3)	(4)	(5)
Female rides (%)	-0.0017^{**}	-0.0017^{**}	-0.0018^{**}	-0.0018^{**}	-0.0016^{**}
	0.0007	0.0007	0.0007	0.0007	0.0007
White riders (%)	0.0008	0.0008	0.0009	0.0010	0.0009
	0.0008	0.0008	0.0008	0.0008	0.0008
Average age of riders	-0.0078^{***}	-0.0079^{***}	-0.0079^{***}	-0.0078^{***}	-0.0080^{***}
	0.0025	0.0025	0.0025	0.0025	0.0025
Number of riders	0.1046^{***}	0.1019^{***}	0.1086^{***}	0.1064^{***}	0.1022^{***}
	0.0301	0.0303	0.0302	0.0304	0.0306
Winnings before VBQ divided by \$100	-0.0240^{***}	-0.0224^{***}	-0.0223^{***}	-0.0211^{***}	-0.0250^{***}
	0.0064	0.0060	0.0061	0.0060	0.0064
Number highly confident and correct answers	0.0229^{**}				0.0021
	0.0106				0.0170
Longest streak of highly confident and correct answers		0.0286^{**}			0.0199
		0.0117			0.0195
Streak of highly confident and correct answers to end game			0.0362^{***}		0.0184
			0.0134		0.0186
Last answer was highly confident and correct				0.1136^{**}	0.0627
				0.0482	0.0660
Constant	0.4993^{***}	0.5228^{***}	0.5841^{***}	0.5546^{***}	0.5148^{***}
	0.1507	0.1476	0.1439	0.1458	0.1498
R^2	0.1082	0.1115	0.1154	0.1101	0.1244

Notes. The dependent variable equals 1 or 0 if the contestant accepted or rejected the gamble of the video bonus question, respectively. Robust standard errors are reported beneath coefficient estimates. There were 379 observations.
***Significant at the 1% level.
**Significant at the 5% level.
*Significant at the 10% level.