RISKY CHOICE IN THE LIMELIGHT

Guido Baltussen, Martijn J. van den Assem, and Dennie van Dolder*

Abstract—This paper examines how risk behavior in the limelight differs from that in anonymity. In two separate experiments, we find that subjects are more risk averse in the limelight. However, risky choices are similarly path dependent in the different treatments. Under both limelight and anonymous laboratory conditions, a simple prospect theory model with a path-dependent reference point provides a better explanation for subjects' behavior than a flexible specification of expected utility theory. In addition, our findings suggest that ambiguity aversion depends on being in the limelight, that passive experience has little effect on risk taking, and that reference points are determined by imperfectly updated expectations.

I. Introduction

INDIVIDUAL decision making is at the core of both economics and psychology. Continuous research efforts have resulted in a rich literature. Still, a persistent concern about empirical research in this field is that specific contextual aspects may restrict the generalizability of findings. Each laboratory or field setting provides its own unique context that cannot be disregarded a priori (Loewenstein, 1999; Levitt & List, 2007a, 2007b; Falk & Heckman, 2009; Camerer, 2015). One particular aspect of the context is the degree of public scrutiny under which a decision is made. Psychological research indicates that the mere presence of others can facilitate performance in simple tasks but impair it in more complex ones (Zajonc, 1965; Bond & Titus, 1983), and that the expectation that one may have to justify one's decisions to observers creates a desire to make decisions that others will judge favorably (Lerner & Tetlock, 1999).

This paper examines to what extent public scrutiny affects risk behavior. In our professional and private lives, we make risky decisions under varying degrees of scrutiny. Mapping the influence of this contextual aspect is therefore an important step in broadening the scope of our understanding of risky choice. Also, from a methodological point of view, it is useful to know to what extent findings on risk preferences from a behavioral laboratory generalize to actual situations with more scrutiny and whether risky choices observed in a high-scrutiny field setting resemble those in a situation with more privacy.

A special example of the relevance of our research question is in the growing literature that studies decision making under risk on the basis of TV game shows. Shows that have been used include *Card Sharks* (Gertner, 1993), *Jeopardy!* (Metrick, 1995), *Illinois Instant Riches* (Hersch & McDougall, 1997), *Lingo* (Beetsma & Schotman, 2001), *Hoosier Millionaire* (Fullenkamp, Tenorio, & Battalio, 2003), *Deal or No Deal* (Post et al., 2008), and *Who Wants to Be a Millionaire?* (Hartley, Lanot, & Walker, 2013). These shows offer unique opportunities to increase our understanding of how individuals and households make consequential risky decisions such as stock market investments and the purchase, insurance, and financing of property.[1] Given the attractive and distinguishing combination of features that game shows can have, more papers based on game shows are likely to appear. Some critics, however, question the external validity of game show research, arguing that contestants' choices might be influenced by pressures from the audience and distress from being in the limelight. As Gertner (1993, p. 519), for example, noted, "If contestants care about the entertainment they provide, they may make riskier decisions than they otherwise would."

Economic studies on public scrutiny are primarily focused on social preferences (Levitt & List, 2007a, 2007b). Surprisingly, whether and how public scrutiny influences risky choice has received relatively little attention from both economists and psychologists. Weigold and Schlenker (1991) find evidence that subjects display a degree of risk tolerance they believe to be judged favorably by observers. Vieider (2009) presents evidence that loss aversion decreases when subjects are made accountable and attributes this to the ease with which his subjects could recognize loss aversion as a bias and their wish to avoid the negative judgments that could result from displaying it. A potentially important issue regarding these two studies is the absence of real incentives, which made it costless for subjects to make a choice that is not truly preferred but thought to be more justifiable in the eyes of onlookers. For hypothetical and incentivized tasks, Miller and Fagley (1991), Takemura (1993, 1994), and Vieider (2011) find that gain and loss framing effects decrease when subjects are made accountable.

First and foremost, this study contributes to the risky choice literature by comparing risky decision making in

Received for publication May 13, 2012. Revision accepted for publication September 24, 2014. Editor: Philippe Aghion.

* Baltussen: Erasmus University Rotterdam; van den Assem: VU University Amsterdam; van Dolder: University of Nottingham.

We are grateful to Han Bleichrodt, Chen Li, Peter Wakker, and the anonymous reviewers for their constructive and valuable comments. We thank Thierry Post for his help in designing and conducting the experiments, Marc Schauten for acting as game show host, and Nick de Heer for his skillful research assistance. The paper benefited from discussions with seminar participants at the Erasmus University of Rotterdam, and with participants of BDRM 2012 Boulder, FUR 2012 Atlanta, M-BEES 2012 Maastricht, SABE 2012 Granada, TIBER 2012 Tilburg, and SPUDM 2013 Barcelona. We gratefully acknowledge support from the Erasmus University of Rotterdam, the Netherlands Organisation for Scientific Research (NWO), the Tinbergen Institute, and the Economic and Social Research Council via the Network for Integrated Behavioral Sciences (ES/K002201/1). Van den Assem and van Dolder conducted much of this research while they were affiliated with the Erasmus University of Rotterdam, the Netherlands.

[1] Game shows have been deployed on various other research domains as well, including strategic decision making (Bennett & Hickman, 1993; Berk, Hughson, & Vandezande, 1996; Tenorio & Cason, 2002), discrimination (Levitt, 2004; Antonovics, Arcidiacono, & Walsh, 2005), bargaining (van Dolder et al., 2015), and cooperative behavior (List, 2004a, 2006; Belot, Bhaskar, & van de Ven, 2010; Oberholzer-Gee, Waldfogel, & White, 2010; van den Assem, van Dolder, & Thaler, 2012).

The Review of Economics and Statistics, May 2016, 98(2): 318–332

doi:10.1162/REST_a_00505

and out of the limelight. It also adds some evidence to the literature on ambiguity aversion by comparing the effect of ambiguity under these two conditions, and to a recent literature on the effect of experience on choices by comparing the behavior of subjects with and without passive experience. Finally, our estimations of structural models of choice add to the literature on reference point formation.

To analyze how risky choice in the limelight differs from that under standard experimental conditions, we conducted two incentivized experiments that mimicked the game of the TV show *Deal or No Deal* (DOND). The next section describes DOND and explains why we used this game. In both experiments, we employed laboratory and limelight treatments. In the laboratory treatments, subjects made decisions in the anonymity of a standard, computerized laboratory setting as typically employed in economic experiments. In the limelight treatments, subjects made their choices in a simulated game show environment, with a live audience, a game show host, and video cameras. With these two conditions—one anonymous and one entailing a high level of public scrutiny—we follow the recommendation of List, Sadoff, and Wagner (2011) to divide experimental samples over the end points of the possible treatment range. By using a game show environment to create public scrutiny, we also shed light on the validity of game shows as natural risky choice experiments.[2]

We consider two ways in which the differences between the treatments can influence risky choice. First, we investigate whether the general degree of risk taking differs between treatments. Second, we examine whether the pattern of path-dependent risk behavior is affected. Earlier DOND-based research found that people show path dependency in the sense that they take more risk if the game develops either substantially worse or substantially better than earlier expectations (Post et al., 2008). These two effects are known, respectively, as the break-even and house-money effect (Kameda & Davis, 1990; Thaler & Johnson, 1990).

If only the general degree of risk taking is affected, this is problematic only insofar as risk preferences are measured in one setting and used to derive point predictions about behavior in another setting. It would imply that it is incorrect to apply the same risk preference parameters across different settings. If, however, the pattern in risky choice is different, the repercussions are potentially more involved because it would mean that we cannot use the same type of risky choice model across different settings.

Our results show that subjects are more risk averse in the limelight than in the anonymity of a typical behavioral laboratory. Simple statistics, probit analyses, and structural choice model estimations consistently lead to this conclusion for both our experiments. The estimates for our structural choice models suggest that the impact of the limelight on risk preference parameters is substantial.

At the same time, however, we observe a similar pattern of path-dependent risk behavior in the laboratory and limelight treatments. Under both experimental conditions, our simple prospect theory–inspired model (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) with a path-dependent reference point provides a better explanation for subjects' behavior than a flexible specification of expected utility theory. Although our study is not designed to point out whether prospect theory or expected utility theory has greater descriptive power—and any conclusion in this direction would depend on the precise empirical implementation of these two theories—it does show that the combination of elements included in our prospect theory model comes closer to the appropriate descriptive model of risky choice and that this finding holds both in and out of the limelight.

Three other noteworthy findings from our analyses are related to ambiguity aversion, the effect of experience, and reference-point formation. First, a design difference between the two sets of experiments that we conducted reveals that the effect of ambiguity depends on being in the limelight or not. Under limelight conditions, subjects take less risk in tasks where they experience some uncertainty about the distribution of possible outcomes than in tasks where the distribution is known. This difference in behavior is absent under laboratory conditions. Second, passive experience does not seem to affect loss aversion or risk aversion in general. One of our experiments featured a comeback treatment with subjects who had seen others perform the experimental task at an earlier occasion. Comparisons between treatments show that their behavior is largely similar to that of inexperienced subjects. Last, we find evidence that preferences are based on imperfectly updated expectations. For all treatments, the parameter estimates of our prospect theory model indicate that subjects' reference points are influenced by their initial beliefs about task outcomes.

The paper proceeds as follows. Section II describes the design, procedure, and results of our first experiment. Section III reports on our second experiment. Section IV discusses our results and concludes.

II. First Experiment

A. Design and Procedure

The experiment followed the basic setup of the popular TV game show *Deal or No Deal*. In DOND, contestants are repeatedly asked to make choices between a sure amount and a risky lottery. At the start, a contestant chooses one (brief)case out of a total of 26 numbered (brief)cases. Each closed case contains one of the game's 26 randomly distributed and widely ranging monetary amounts. After selecting

[2] For domains other than risky choice, a number of studies have investigated this issue before. Tenorio and Cason (2002), Healy and Noussair (2004), and Antonovics, Arcidiacono, and Walsh (2009) observe how students play *The Price Is Right* and *The Weakest Link* under laboratory conditions and find that their behavior or performance is similar to that of contestants in the TV show.

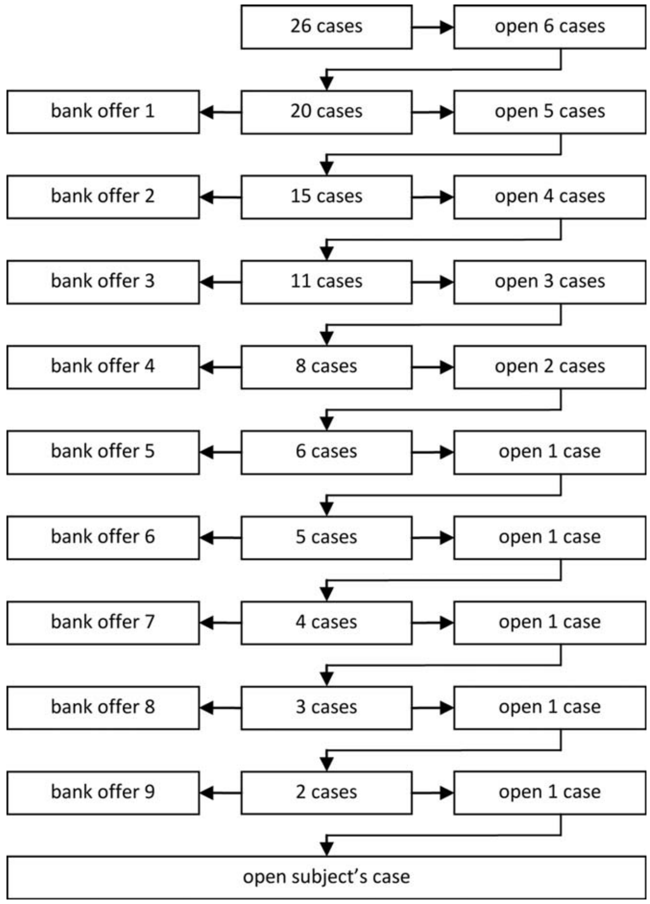
this personal case, a contestant has to select six of the other cases to be opened. The prizes in these cases are revealed and no longer in play, thereby increasing the information on the prize in the contestant's personal case. After the contents of six cases have been revealed, an imaginary "banker" offers to buy the contestant's case. If the contestant decides "Deal," she receives the amount offered, and the game ends. If the contestant decides "No deal," the game continues and she has to open five additional cases. Based on the then remaining set of 15 prizes, the banker makes a new offer. The contestant again has to decide either "Deal" or "No deal." After a "No deal," this process continues until the contestant accepts an offer, or until no case other than the contestant's own case is left and she receives the content of this case. The bank offer typically starts as a small percentage of the average remaining prize, and this percentage gradually increases as the game proceeds. The game lasts for a maximum of nine rounds. The number of cases to be opened in each round is 6, 5, 4, 3, 2, 1, 1, 1, and 1, reducing the number of remaining cases from 26 to 20, 15, 11, 8, 6, 5, 4, 3, 2, and finally 1. Figure 1 presents a schematic overview of the course of the game.

In the experiment, subjects played DOND for real incentives in either a computer laboratory (laboratory treatment) or a classroom mimicking a TV studio (limelight treatment). The prizes in the experiment were equal to the prizes used in the original Dutch edition of the TV show, scaled down by a factor of 10,000. The smallest amounts were rounded up to 1 cent. The resulting set of prizes was €0.01 (nine times due to rounding up); €0.05; €0.10; €0.25; €0.50; €0.75; €1; €2.50; €5; €7.50; €10; €20; €30; €40; €50; €100; €250; €500. The distribution of the prizes was clearly positively skewed, with a median of €0.63 and a mean of €39.14. Figure 2 demonstrates how the game was shown to subjects.

The laboratory treatment was conducted as a typical economic experiment. Subjects played DOND in the quiet, controlled environment of a computerized laboratory and made their choices on a private computer terminal. The setting was designed to minimize potential scrutiny from other subjects. In particular, computers surrounding a given subject were empty, and a sunken screen and dividers were used to ensure privacy.

The limelight treatment was designed to replicate a TV studio as closely as possible. The experiment took place in a theater-style lecture room. Subjects made their decisions on a lighted stage in front of a live audience consisting of fellow students and some university employees. They were guided through the experiment by a game show host, played by a popular lecturer. Furthermore, video cameras were pointed at the subject on stage. The game was shown on a computer monitor in front of the subject and projected on a large screen for the audience. Members of the audience were allowed to applaud, shout hints, and the like. Before a game started, the host had a brief introductory talk with the subject on stage, covering basic topics such as the subject's name, age, favorite sports, and other interests.

Figure 1.—Flowchart of the Game



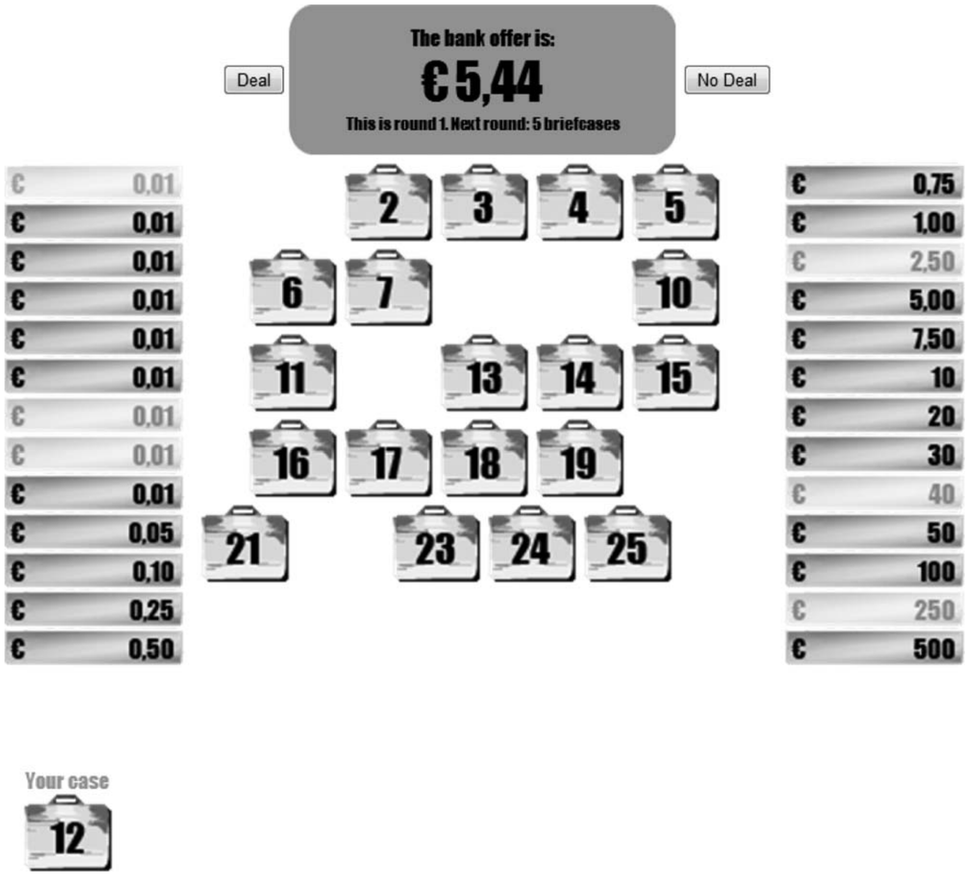
In each of a maximum of nine game rounds, the subject chooses a given number of cases to be opened. After the prizes in the chosen cases are revealed, an imaginary banker offers to buy the subject's own case. If the subject accepts the offer ("Deal"), she receives the amount offered and the game ends. If the subject rejects the offer ("No deal"), play continues and she enters the next round. If the subject decides "No deal" in the ninth round, she receives the prize in her own case (from Post et al., 2008.)

The data from the limelight treatment have previously been analyzed in Post et al. (2008).[3] To facilitate comparisons with the actual game show in that study, each subject replayed one of the first forty scenarios from the Dutch version of DOND: independent of the order in which a subject opened the numbered cases, the order in which the prizes and the offers appeared corresponded exactly to the original scenario. In addition, we matched the gender of subjects and TV contestants: female (male) subjects were randomly assigned to scenarios from female (male) contestants. We did not select these forty scenarios to encourage or avoid particular situations or behaviors. Rather, subjects played games that had been randomly generated earlier.[4] The instructions were as similar as possible to those that had

[3] The limelight treatment was employed there to analyze the isolated effect of the amounts at stake. (Another treatment was conducted under identical limelight conditions but used stakes that were a factor of ten larger.)

[4] If a subject played on longer than the original contestant, we had no information on eliminated prizes and bank offers from that point onward. We then randomly selected the eliminated prizes ourselves (holding them constant across treatments) and set the offers according to the pattern observed for the TV episodes.

FIGURE 2.—EXAMPLE OF THE GAME AS DISPLAYED ON THE EXPERIMENTAL SCREENS



The various prizes are listed in the columns on the left and right sides. Prizes that are eliminated are blurred. The current bank offer is shown at the top, and the subject or host can select either "Deal" or "No deal" by clicking on the respective button. The remaining cases are shown in the center of the screen, and the subject's own case is in the bottom left-hand part. This example shows the two options open to a subject after opening six cases in the first round: accept a bank offer of €5.44 or continue to play with the remaining twenty cases. Note that a comma is used to separate decimals here, as this is common for our subjects.

been handed out to TV show contestants. Subjects received the original Dutch instructions used for the TV version, plus a cover sheet explaining the experiment. We did not impose any time constraint.

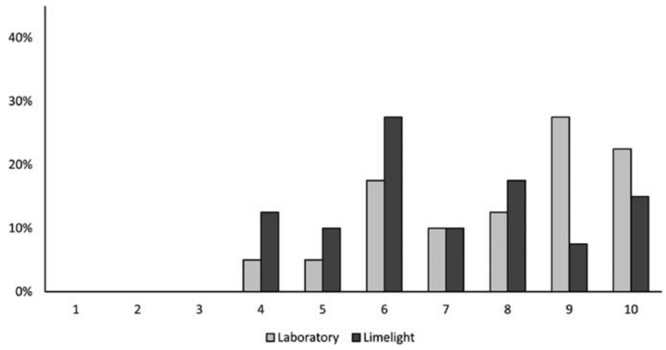
The subjects were randomly selected from a larger population of business or economics students at the Erasmus University of Rotterdam who had applied to participate in economic experiments. Forty subjects took part in the limelight treatment, and forty took part in the laboratory treatment: one for each of the forty scenarios in both treatments. We subdivided subjects in the limelight treatment across two separate sessions. In total, eighty students were invited to the two limelight sessions—approximately forty per session. This was done to ensure a sufficiently large audience and create a buffer in case some subjects did not show up. After one subject had finished playing the game, a new subject was selected to play, until twenty subjects had played the game. Hence, approximately half of the students in each session were selected to play. Subjects were paid according to the outcome of their game. Subjects who were not selected received no pay. Each game lasted about 5 to 10 minutes, and an entire session lasted approximately 2.5 hours. The forty subjects who were selected for the laboratory treatment were similarly subdivided across two different sessions. In each session, twenty subjects played the game simultaneously.

Using the game of DOND has several benefits. Its appealing qualities have attracted considerable research attention, making it the most frequently studied game show in the domain of risky choice (Blavatskyy & Pogrebna, 2010; Brooks et al., 2009a, 2009b; Deck, Lee, & Reyes, 2008; Post et al., 2008). The game involves only simple stop-go decisions ("Deal" or "No deal") that require no or minimal skill, knowledge, or strategy. Moreover, the dynamic nature of the game allows not only comparing general levels of risk taking between treatments but also the pattern of path dependence. In addition, subjects may find it relatively natural to make decisions in front of an audience when the task at hand is from a TV game show, and the entertainment value of DOND may help to involve the audience in the game. The great popularity of the game on TV brings the advantage that it is generally well understood by subjects.

B. *Descriptive Statistics and Probit Analysis*

We observed 579 decisions made by eighty subjects. A crude way to investigate differences in risky choice

FIGURE 3.—DISTRIBUTIONS OF STOP ROUNDS (FIRST EXPERIMENT)



The figure depicts the distribution of the stop round for the two treatments of our first experiment. The stop round is the round in which the bank offer is accepted ("Deal"), or ten for subjects who rejected all offers. In the laboratory treatment, subjects played the game in a standard economic laboratory setting, while in the limelight treatment, subjects played the game in an environment mimicking a TV studio with a live audience.

TABLE 1.—PROBIT REGRESSION RESULTS: FIRST EXPERIMENT

	Coefficient
Constant	−1.340 (0.036)
EV/100	1.836 (0.000)
EV/*BO*	−1.188 (0.004)
Stdev/*EV*	2.186 (0.000)
Limelight	0.509 (0.004)
LL	−131.1
McFadden R^2	0.355
No. obs.	579

The table displays the maximum likelihood estimation results of a probit model aimed at explaining the decisions of the subjects in the laboratory ($N = 40$) and limelight ($N = 40$) treatment of our first experiment. The dependent variable is the subject's decision, with a value of 1 for "Deal" and 0 for "No deal." *EV* is the current average remaining prize in euros, *BO* is the bank offer in euros, *Stdev* is the standard deviation of the distribution of the average remaining prize in the next game round, and *Limelight* is a dummy variable that takes a value of 1 for observations from the limelight treatment. In addition to the maximum likelihood estimates for the regression coefficients, the table reports the log likelihood (LL), McFadden R^2, and the total No. obs. The *p*-values (within parentheses) are corrected for correlation between the responses of a given subject (subject-level cluster correction).

between the treatments is to compare subjects' stop rounds. The stop round is the round in which a subject decides to accept the bank offer ("Deal"), or 10 if she rejects all nine offers. As the bank offer gradually increases as a percentage of the average remaining prize, deciding "Deal" at a relatively early (late) stage implies a relatively high (low) degree of risk aversion.

Figure 3 shows the distribution of the stop round for both treatments. Subjects in the limelight treatment decide "Deal" earlier than subjects in the laboratory treatment. The average stop round in the limelight treatment is 6.93 compared to 7.93 in the laboratory treatment. The difference of exactly one round is statistically significant (*t*-test: $p = 0.019$; Mann-Whitney *U* test: $p = 0.021$).

The stop round is a crude measure because it does not reflect differences in the actual bank offer, the stakes, or the risk of continuing play. To control for these factors, we perform a probit regression analysis. The dependent variable is the subject's decision, taking the value of 1 for "Deal" and 0 for "No deal." We explain subjects' decisions using the following variables:

- *EV*/100: Included to control for the stakes and calculated as the current average remaining prize (divided by €100 for more convenient regression coefficients)
- *EV*/*BO*: Included to control for the expected return of continuing play and calculated as the average remaining prize divided by the bank offer, or the expected relative return (+1) from rejecting both the current and all subsequent bank offers
- *Stdev*/*EV*: Included to control for the riskiness of continuing play and calculated by dividing the standard deviation of the distribution of the average remaining prize in the next round by the current average remaining prize
- *Limelight*: The main variable of interest, a dummy variable that takes the value of 1 if the choice was made in the limelight treatment and 0 if it was made in the lab treatment

We do not consider the common demographic characteristics age and gender. Our subjects are all students of about the same age, and gender does not have significant explanatory power. We allow the possibility that errors of individual subjects are correlated through cluster corrections on the standard errors (Wooldridge, 2003).

Table 1 shows the regression results. As expected for risk-averse individuals, the propensity to "Deal" is positively related to the riskiness of continuing play, and negatively related to the expected return of continuing play Furthermore, the "Deal" propensity increases with the stakes. Consistent with the simple analysis of stop rounds, subjects in the limelight are more likely to "Deal" than those in the laboratory ($p = 0.004$).

In the context of DOND, people have been shown to take more risk after earlier expectations have been shattered or surpassed. In order to investigate this pattern descriptively for our two treatments, we classify subjects as being "losers," "neutrals," or "winners." We follow the method of Post et al. (2008), which takes into account the downside risk and the upside potential of rejecting a bank offer. In particular, we define a subject's best-case scenario (BC_r) and worst-case scenario (WC_r) of opening another case in round r as

$$BC_r = \frac{n_r \bar{x}_r - x_r^{\min}}{n_r - 1}, \tag{1}$$

$$WC_r = \frac{n_r \bar{x}_r - x_r^{\max}}{n_r - 1}, \tag{2}$$

where n_r is the number of remaining cases in round r, $\bar{x}_r$ is the average remaining prize in round r, and $x_r^{\min}$ and $x_r^{\max}$ stand for the smallest and largest remaining prize, respectively. A subject is classified as a loser if her BC_r belongs to the worst one-third of all subjects in that round and as a winner if her WC_r belongs to the best one-third. Game situations that satisfy neither condition (or both) are classified as neutral. If two subjects share the same BC_r or WC_r

TABLE 2.—DECISIONS AFTER BAD AND GOOD FORTUNE: FIRST EXPERIMENT

Round	Loser			Neutral			Winner		
	%BO	No.	%D	%BO	No.	%D	%BO	No.	%D
A. Laboratory									
1	6	14	0	6	12	0	6	14	0
2	15	14	0	13	12	0	15	14	0
3	42	13	0	32	14	0	32	13	0
4	68	14	0	61	12	8	56	14	7
5	83	13	0	74	12	0	73	13	15
6	92	12	8	88	12	25	86	12	25
7	98	9	33	99	11	0	94	9	11
8	104	6	17	101	13	23	102	6	17
9	101	5	20	102	10	70	104	5	60
1–9		100	6		108	13		100	11
B. Limelight									
1	6	14	0	6	12	0	6	14	0
2	15	14	0	13	12	0	15	14	0
3	42	13	0	32	14	0	32	13	0
4	68	14	0	61	12	25	56	14	14
5	81	12	0	76	11	18	74	12	17
6	92	11	9	87	9	44	88	11	55
7	94	4	25	98	12	25	93	4	0
8	106	4	0	101	8	50	102	4	75
9	108	1	0	101	7	29	105	1	100
1–9		87	2		97	19		87	16

The table summarizes the decisions of the subjects in the laboratory (panel A: $N = 40$) and limelight (panel B: $N = 40$) treatment of our first experiment. The samples are split based on the fortune experienced during the game. A subject is classified as a "loser" ("winner") if her average remaining prize, after eliminating the lowest (highest) remaining prize, is among the worst (best) one-third for all subjects in the same game round. The table displays the percentage bank offer (%BO), the number of subjects (No.), and the percentage of subjects choosing "Deal" (%D) for each category and game round.

but one falls below the one-third cutoff and one above it, then both are classified as neutral.

Table 2 shows the choices of subjects conditional on the classification of their game situation. In both treatments, winners and losers continue play more often than subjects in the neutral group. This difference is especially pronounced for losers.

C. *Structural Models*

We now move to the estimation of structural choice models in order to examine how the more risk-averse behavior in the limelight as opposed to the laboratory corresponds to differences in risk preference parameters and to further investigate the pattern of path dependence. We implement two simple structural models: one in the spirit of expected utility theory (EU) and the other inspired by prospect theory (PT).[5]

Structural choice models allow a wide range of specifications. For example, there are many ways to specify the utility function, the error term, reference point dynamics, and probability weighting. We follow the methodology used in the earlier DOND-based studies by Post et al. (2008) and Baltussen et al. (2012) and summarize this approach below. For further methodological details, background and discussion, we refer to these two prior studies.

For our EU specification, we apply a flexible-form expo-power utility function that allows the combination of increasing relative risk aversion (IRRA) and decreasing absolute risk aversion (DARA):

$$u(x) = \frac{1 - \exp(-\alpha x^{1-\beta})}{\alpha}, \tag{3}$$

where α and β are the risk aversion coefficients, subject to $\alpha\beta \geq 0$ to exclude (more exotic) utility functions that combine concavity and convexity.[6] The expo-power function reduces to a CRRA (constant relative risk aversion) power function when $\alpha \to 0$ and to a CARA (constant absolute risk aversion) exponential function when $\beta = 0$.

For PT, we use a simple representation that incorporates loss aversion, uses probabilities as decision weights, and has equal curvature for gains and losses. In particular, the value function is defined as

$$v(x|RP) = \begin{cases} (x - RP)^{\alpha} & x > RP \\ -\lambda(RP - x)^{\alpha} & x \leq RP \end{cases}, \tag{4}$$

where $\lambda > 0$ is the loss-aversion parameter, RP is the reference point, and $\alpha > 0$ represents the curvature of the value function.

[5] Henceforth, we refer to these models as the EU model and the PT model. We acknowledge that both theories can be implemented through numerous different and sometimes overlapping specifications. The fit for EU could, for example, be improved with an even more flexible utility function that has both concave and convex segments. As we explained in Section I, our study does not aim to point out whether prospect theory or expected utility theory has greater descriptive power.

[6] Post et al. (2008) and Baltussen et al. (2012) impose $\beta \geq 0$. The present specification gives the EU model greater flexibility to capture risk-seeking behavior. Using $\beta \geq 0$ instead of $\alpha\beta \geq 0$ reduces the expo-power function to a CARA function ($\beta = 0$) for all treatments and yields similar treatment effects. Also, we do not follow these two studies in including initial wealth as a free parameter. This simplification is in line with the standard approach in experimental research and rules out the possibility of erroneously capturing differences in risk aversion between randomized treatments by differences in wealth estimates.

Recent literature suggests that reference points are expectation based and dynamically but partially updated (Kőszegi & Rabin, 2006, 2007; Abeler et al., 2011; Baucells, Weber, & Welfens, 2011; Ericson & Fuster, 2011). In this spirit, the reference point in round r, RP_r, is modeled as a function of the current bank offer, B_r, and the relative increase of the average remaining prize during the game, $d_r = (\bar{x}_r - \bar{x}_0)/\bar{x}_r$,

$$RP_r = (\theta_1 + \theta_2 d_r)\, B_r, \tag{5}$$

where $\theta_1 < 1$ ($\theta_1 > 1$) indicates that the reference point generally takes a value below (above) the current bank offer and $\theta_2 \leq 0$ allows for (imperfect) updating of the reference point across rounds. $\theta_2 = 0$ reflects perfect updating, while $\theta_2 < 0$ implies that the reference point sticks to initial expectations. To illustrate, when $\theta_1 = 1$ and $\theta_2 = 0$, the reference point equals the current bank offer; when $\theta_1 = 1$ and $\theta_2 = -1$, the reference point corresponds to the amount that would have been on offer if the average prize had been at its starting level; when $\theta_1 = 0$ and $\theta_2 = 0$, the reference point is 0 and all outcomes are considered gains. Combined with loss aversion and a value function that is concave for gains and convex for losses, the reference point model allows break-even and house-money effects.

Post et al. (2008) and Baltussen et al. (2012) also include a separate term for changes during the last two rounds. We drop this short-term lag for brevity and convenience. Baltussen et al. (2012) also found that intermediate changes are economically and statistically insignificant for the reference point. Including the term has no material effect on the other parameters for each of our treatments.[7] Moreover, the use of one single stickiness parameter facilitates comparisons between treatments.

We make the simplifying assumption that subjects look ahead only one round, implying that they compare the current bank offer with the distribution of possible bank offers in the next round. As Post et al. (2008) explained, assuming a myopic frame rather than multistage backward induction is behaviorally plausible and does not materially affect the results. Post et al. (2008) also show that the percentage bank offer can be adequately captured by the simple function

$$b_{r+1} = b_r + (1 - b_r)\rho^{(9-r)}, \tag{6}$$

where b_r is the percentage bank offer relative to the expected value of the remaining prizes in round r and ρ measures the speed at which it approaches the expected value ($0 \leq \rho \leq 1$). Post et al. (2008) estimate ρ to be 0.832 for the forty episodes of the Dutch edition of DOND that we used as scenarios in our experiment. In our analysis, we treat this bank offer model as deterministic and known to the subjects.

[7] The unimportance of recent changes for the reference point in experiments can be explained by the shorter duration of a game. The original model was designed to capture the behavior of contestants in the TV version, where the recording of a game lasts for about 1 hour and where recent developments are thus more salient. In our experiments, a game lasts no more than 10 minutes, increasing the likelihood that subjects simply compare their current situation with that at the start of their game.

Table 3.—Expected Utility Model Estimates: First Experiment

	Laboratory	Limelight
α	—	0.021 (0.000)
β	−0.861 (0.022)	0.000 (1.000)
σ	0.544 (0.000)	0.332 (0.000)
LL	−85.5	−78.7
AIC	177.0	163.4
BIC	188.2	174.2
No. obs.	308	271
CC (0/1)	1.378	0.995
CC (0/10)	1.378	0.948
CC (0/100)	1.378	0.554
CC (0/1000)	1.378	0.067

The table displays the maximum likelihood estimation results of our EU model for the laboratory (panel A: $N = 40$) and limelight (panel B: $N = 40$) treatment of our first experiment. Shown are the risk aversion parameters (α and β) of the utility function and the noise parameter (σ). The table also shows the log likelihood (LL), the AIC and BIC statistics, and the number of decisions (No. obs.). The implied certainty coefficient (CC; certainty equivalent as a fraction of the expected value) is shown for 50/50 gambles of €0 or €10^z, $z = 0, 1, 2, 3$. The p-values (within parentheses) are corrected for correlation between the responses of a given subject (subject-level cluster correction).

We apply maximum likelihood techniques to estimate the unknown parameters. The likelihood of each decision is based on the utility difference between the current bank offer and future bank offers. We assume that a decision is more difficult if the standard deviation of the utility values from continuing play is larger and set the standard deviation of the model error proportional to this measure. To reduce the potential influence of individual observations, we truncate the likelihood of each decision at a minimum of 1%.

Table 3 gives the results of the EU model. For the laboratory treatment, the expo-power function converges to a risk-seeking CRRA power function. In terms of explanatory power, this model outperforms a naive model that assumes risk neutrality ($\chi^2(2) = 24.27$, $p < 0.001$). In contrast, the function reduces to a risk-averse CARA exponential function for the limelight treatment. This model also fits the data better than a risk-neutral model ($\chi^2(2) = 10.29, p = 0.006$).

The shapes of the estimated utility functions are thus very different for the two treatments: one is convex and the other concave. Certainty equivalents (CEs) and certainty coefficients (CCs) can help to interpret the degrees of risk aversion implied by the models. The values nicely illustrate the substantial differences between the two treatments. For a lottery with a 50% chance of €100 and €0 otherwise, the implied CE under limelight conditions is €27.72. The CC is 27.72 / 50.00, or 55%. For the laboratory treatment, the CE (CC) of €68.91 (138%) is well above the expected value (100%).

The EU specification has difficulties capturing the different preferences of losers, neutrals, and winners (as defined earlier). This is illustrated in table 4, which reports separate EU-model estimates for the subsamples. In the limelight, the estimated utility function for losers reflects a preference for risk, while neutrals and winners are risk averse. In the laboratory, each subgroup is best described by a model of risk-seeking preferences, but losers are more risk prone than neutrals and winners. The CCs illustrate the differences between the utility functions.

TABLE 4.—PATH DEPENDENCE: FIRST EXPERIMENT

	Loser	Neutral	Winner
A. Laboratory			
α	—	—	—
β	−1.459 (0.070)	−0.459 (0.290)	−0.668 (0.101)
σ	0.530 (0.000)	0.431 (0.007)	0.577 (0.000)
LL	−22.0	−32.9	−28.5
No. obs.	100	108	100
CC (0/1)	1.509	1.244	1.320
CC (0/10)	1.509	1.244	1.320
CC (0/100)	1.509	1.244	1.320
CC (0/1000)	1.509	1.244	1.320
B. Limelight			
α	−2.251 (0.095)	0.018 (0.064)	0.027 (0.000)
β	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)
σ	0.271 (0.000)	0.374 (0.000)	0.252 (0.000)
LL	−8.3	−34.5	−20.2
No. obs.	87	97	87
CC (0/1)	1.473	0.995	0.993
CC (0/10)	1.938	0.954	0.932
CC (0/100)	1.994	0.595	0.464
CC (0/1000)	>1.999	0.076	0.051

The table shows the maximum likelihood estimation results of our EU model for subsamples from the laboratory (panel A) and limelight (panel B) treatment of our first experiment. For each treatment, the sample is split based on the fortune experienced during the game. A subject is classified as a "loser" ("winner") if her average remaining prize, after eliminating the lowest (highest) remaining prize, is among the worst (best) one-third for all subjects in the same game round. Definitions are as in table 3.

TABLE 5.—PROSPECT THEORY MODEL ESTIMATES: FIRST EXPERIMENT

	Laboratory	Limelight
α	0.554 (0.000)	0.711 (0.001)
λ	1.505 (0.042)	2.825 (0.000)
θ_1	1.014 (0.000)	1.040 (0.000)
θ_2	−0.045 (0.001)	−0.072 (0.019)
σ	0.334 (0.000)	0.257 (0.000)
LL	−66.8	−63.6
AIC	143.7	137.2
BIC	162.3	155.2
No. obs.	308	271
CC (0%)	0.572	0.754
CC (100%)	0.960	0.796
CC (200%)	1.428	1.246

The table displays the maximum likelihood estimation results of our PT model for the laboratory (panel A: $N = 40$) and limelight (panel B: $N = 40$) treatment of our first experiment. Shown are the loss aversion (λ) and curvature (α) parameters of the value function, the two parameters of the reference point model (θ_1 and θ_2), and the noise parameter (σ). The table also shows the log likelihood (LL), the AIC and BIC statistics, and the number of decisions (No. obs.). The implied certainty coefficient (CC; certainty equivalent as a fraction of the expected value) is shown for 50/50 gambles of €0 or €10^z, for any $z > 0$, assuming that the reference point equals 0%, 100%, or 200% of the expected value. For λ and α, the null hypotheses are that these parameters equal unity, implying no utility curvature and no loss aversion. The other parameters are tested relative to zero. The p-values (within parentheses) are corrected for correlation between the responses of a given subject (subject-level cluster correction).

Table 5 shows the PT estimates. In the laboratory treatment, we find a rather strong utility curvature, with an α of 0.554. The loss aversion coefficient, λ, equals 1.505. Both values differ significantly from unity (α: $p < 0.001$; λ: $p = 0.042$). Furthermore, the reference point sticks to earlier expectations, with $\theta_2 = -0.045$ ($p = 0.001$). In the absence of changed expectations, it takes a value that is close to the current bank offer ($\theta_1 = 1.014$).

In the limelight treatment, utility curvature ($\alpha = 0.711$, $p = 0.001$) and loss aversion ($\lambda = 2.825$, $p < 0.001$) occur as well. Again, the reference point is sticky ($\theta_2 = -0.072$, $p = 0.019$) and, on average, close to the bank offer ($\theta_1 = 1.040$). While the curvature and reference point parameters are not significantly different between the two treatments, loss aversion is stronger in the limelight than in the laboratory (α: $p = 0.142$; λ: $p = 0.003$; θ_1: $p = 0.166$; θ_2: $p = 0.423$).

Finally, note that the PT model, which can capture the path dependence of risk attitudes through a sticky reference point, loss aversion, and reflection of the value function around the reference point, explains subjects' choices significantly better than the EU model. This better fit holds for both the limelight treatment and the laboratory treatment, and when we account for the larger number of parameters as compared to the EU model (consider the very different AIC and BIC values).

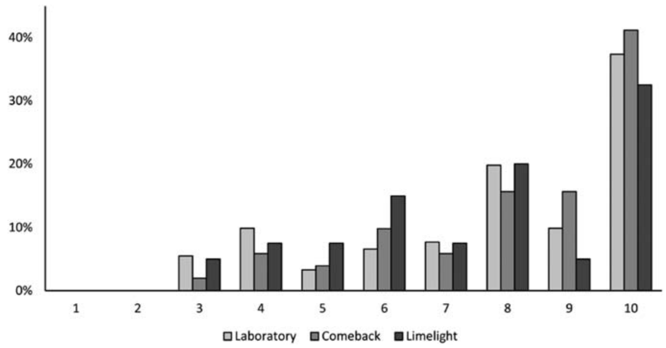
III. Second Experiment

A. Design and Procedure

To investigate the robustness of the results, we conducted a second experiment. Below we list the design differences. In all other respects, the new experiment was identical to the previous one.

First, we used fixed percentage bank offers. Although subjects in the first experiment had been informed about the two most important factors that determine the bank offer (the bank offer strongly depends on the average remaining prize and the percentage bank offer gradually increases over the rounds), subjects still faced some ambiguity about the precise offers. Consequently, we cannot exclude that the treatment effects are related to ambiguity rather than risk preferences (Ellsberg, 1961; Camerer & Weber, 1992). In the second experiment, we therefore used fixed percentage bank offers for each game round. That is, the bank offer

Figure 4.—Distributions of Stop Round: Second Experiment



The figure depicts the distribution of the stop round for the three treatments of our second experiment. In the comeback treatment, subjects played the game in a standard economic laboratory setting after viewing others play the game in the limelight treatment. Other definitions are as in figure 3.

was a percentage of the expected value of the prize in the subject's case that depended on the round number only. For rounds 1 to 9, the percentages were 15, 30, 45, 60, 70, 80, 90, 100, and 100, respectively. Subjects were informed about this precise structure in the instructions.

Second, we added a third treatment. Subjects in the limelight treatment passively gained experience in playing the game by watching the decisions and outcomes of others. In the first experiment, a subject in the limelight, on average, had watched 9.5 others play the game before she was selected to play herself. In contrast, laboratory subjects did not observe any other subject playing prior to their own game. To examine whether differences in such passive experience matter, we also ran a comeback treatment. This treatment consisted of subjects who had been audience members in the limelight treatment but had not been selected to play the game onstage themselves. These subjects were invited to play the game in the laboratory afterward.

An additional benefit of this approach is that subjects in the limelight treatment now always had the opportunity to play the game. In the first experiment, this was not the case, as those who were not selected went home empty-handed. As a result, a sense of relief or feelings of luck may have influenced the behavior of those selected. Our announcement of the comeback session avoids this possible confound.

Finally, we used completely random scenarios and more formal experimental instructions. Because comparison with the actual game show was not one of the objectives of this new experiment, there was no need to replay scenarios from the original TV show or use the instructions that had been handed out to TV show contestants.

The subjects were a randomly selected subset of first-year economics students at the Erasmus University of Rotterdam who were required to participate as part of a first-year introductory economics course. We observed 91 subjects in the laboratory treatment, 40 in the limelight treatment, and 51 in the comeback treatment.[8] All subjects in the comeback treatment had previously watched 20 subjects play the game onstage in our limelight treatment.

[8] The laboratory treatment of this second experiment is used in Baltussen et al. (2012) to investigate different types of incentive systems. The experiment as a whole also consisted of the other treatments analyzed in Baltussen et al.

Table 6.—Probit Regression Results: Second Experiment

	Coefficient
Constant	−1.519 (0.000)
EV/100	1.090 (0.000)
EV/*BO*	−0.661 (0.000)
Stdev/*EV*	1.267 (0.000)
Limelight	0.293 (0.037)
Comeback	−0.111 (0.414)
LL	−291.1
McFadden R^2	0.258
No. obs.	1367

The table displays the maximum likelihood estimation results of a probit model aimed at explaining the decisions of the subjects in the laboratory ($N = 91$), limelight ($N = 40$), and comeback ($N = 51$) treatment of our second experiment. *Comeback* is a dummy variable that takes a value of 1 for observations from the comeback treatment. Other definitions are as in table 1.

B. Analyses

As with the first experiment, we start with an analysis of the stop rounds.[9] Recall that deciding "Deal" relatively early (late) indicates a relatively high (low) degree of risk aversion. The treatment differences are less pronounced than before. The average stop round in the limelight treatment is 7.55 compared to 7.87 in the laboratory treatment and 8.27 in the comeback treatment. While the average stop round is thus lowest in the limelight treatment, the differences with the two other treatments are not statistically significant (versus laboratory: *t*-test $p = 0.463$, Mann-Whitney U test $p = 0.406$; versus comeback: *t*-test $p = 0.109$, Mann-Whitney U test $p = 0.126$). The difference between the laboratory and the comeback treatment is not significant either (*t*-test: $p = 0.291$; Mann-Whitney U test: $p = 0.362$). Figure 4 shows the distribution of the stop round for the three treatments.

The absence of a treatment effect in the stop rounds may be related to the crudeness of this analysis. While the forty subjects in each of the two treatments in the first experiment played the same forty scenarios as TV show contestants, subjects in this experiment faced completely random scenarios. Decision problems can thus be markedly different among treatments, making it even more important to control for differences in the stakes, bank offer, and risk of continuing play. Therefore, we now move to the probit and structural model analyses. For background on the methods, we refer to the previous section.

Table 6 shows the results of the probit regression. The results closely resemble those of the first experiment. The propensity to "Deal" is positively related to the riskiness of continuing play and to the stakes and negatively to the expected return of continuing play. After controlling for

[9] Three subjects in this experiment ended up with trivial choice problems involving prizes of 1 cent only. Each rejected all nine offers, implying stop round values of 10. The results are not materially different when we set their stop round equal to the number of the first round that had no prizes other than prizes of 1 cent (or to the average of this number and 10). We omit these uninformative choices in the subsequent probit regression analyses and structural choice model estimations.

TABLE 7.—EU AND PT MODEL ESTIMATES: SECOND EXPERIMENT

	Laboratory	Comeback	Limelight
A. Expected Utility Theory			
α	—	—	0.010 (0.000)
β	−0.504 (0.000)	−0.332 (0.006)	0.000 (1.000)
σ	0.375 (0.000)	0.296 (0.000)	0.196 (0.000)
LL	−165.2	−88.1	−71.3
AIC	336.5	182.2	148.6
BIC	350.0	194.1	159.6
No. obs.	677	401	289
CC (0/1)	1.261	1.189	0.998
CC (0/10)	1.261	1.189	0.975
CC (0/100)	1.261	1.189	0.761
CC (0/1000)	1.261	1.189	0.140
B. Prospect Theory			
α	0.408 (0.000)	0.639 (0.000)	0.751 (0.000)
λ	1.259 (0.005)	1.407 (0.000)	1.863 (0.000)
θ_1	1.002 (0.000)	1.015 (0.000)	1.088 (0.000)
θ_2	−0.009 (0.000)	−0.067 (0.005)	−0.154 (0.000)
σ	0.223 (0.000)	0.163 (0.000)	0.231 (0.000)
LL	−142.7	−59.9	−61.0
AIC	295.4	129.8	131.9
BIC	317.9	149.8	150.3
No. obs.	677	401	289
CC (0%)	0.367	0.676	0.794
CC (100%)	0.996	0.951	0.857
CC (200%)	1.633	1.324	1.206

The table displays the maximum likelihood estimation results of the EU (panel A) and PT (panel B) model for the laboratory ($N = 91$), limelight ($N = 40$), and comeback ($N = 51$) treatment of our second experiment. The panels follow the format and definitions of tables 3 and 5.

these variables, subjects in the limelight turn out to be significantly more likely to "Deal" than those in the laboratory ($p = 0.037$) and the comeback treatment ($p = 0.004$). There is no significant difference between the laboratory and comeback treatment ($p = 0.414$).

Panel A of table 7 presents the results of the structural model estimations for EU. As in the previous experiment, the expo-power function converges to a risk-seeking CRRA power function for the laboratory treatment. The same is found for the comeback treatment. In both cases, the estimated model outperforms a naive model that assumes risk neutrality (laboratory: $\chi^2(2) = 30.23$, $p < 0.001$; comeback: $\chi^2(2) = 15.20$, $p < 0.001$). The β parameter is not significantly different between the laboratory and comeback treatment ($p = 0.359$). For the limelight treatment, the expo-power function again reduces to a risk-averse CARA exponential function that outperforms risk neutrality ($\chi^2(2) = 11.97, p = 0.003$).

Panel B of table 7 presents the estimation results for PT. In the laboratory treatment, we find a rather strong utility curvature, with an α of 0.408. Loss aversion is limited, with λ equaling 1.259. Both values differ significantly from unity (α: $p < 0.001$; λ: $p = 0.005$). Furthermore, the reference point sticks to earlier expectations, with $\theta_2 = -0.009$ ($p < 0.001$), and is, on average, located in the vicinity of the bank offer ($\theta_1 = 1.002$). For subjects in the comeback treatment, we find a curvature of 0.639, a loss aversion of 1.407, and reference point parameters of −0.067 and 1.015 (all $p \leq 0.005$). When we compare the various parameters of these two treatments, we find that subjects in the comeback treatment demonstrate less curvature and a stickier and more elevated reference point than subjects in the laboratory treatment (α: $p = 0.001$; λ: $p = 0.278$; θ_1: $p = 0.014$; θ_2: $p = 0.015$).

In the limelight treatment, we similarly find significant values for utility curvature ($\alpha = 0.751$, $p < 0.001$), loss aversion ($\lambda = 1.863$, $p < 0.001$), and stickiness of the reference point ($\theta_2 = -0.154$, $p < 0.001$). In the absence of changed expectations, the reference point takes a value that is relatively close to the current bank offer ($\theta_1 = 1.088$). In line with the first experiment, subjects in the limelight are more loss averse ($p = 0.006$) than subjects in the laboratory. In addition they now also display significantly less curvature of the value function ($p < 0.001$) and a stickier and more elevated reference point (both $p < 0.001$). Compared to the comeback treatment, subjects in the limelight are again more loss averse ($p = 0.040$), and they also have a stickier ($p = 0.003$) and more elevated ($p < 0.001$) reference point. Utility curvature is not significantly different ($p = 0.173$).

Similar to the first experiment, the PT model explains subjects' choices significantly better than the EU model. This better fit holds for all three treatments. Further analyses yield evidence of the same pattern of path dependence as in the first experiment: in all treatments, we find that losers have a greater risk appetite than winners and neutral subjects.

The major difference between the two experiments was that the future bank offers were somewhat ambiguous to subjects in the first and fixed and known to them in the second. A comparison of the estimation results can thus give an indication of whether the effect of ambiguity on behavior is similar or different in and out of the limelight.[10]

Because the bank offer structure differs between the experiments and the stop round and probit analyses cannot take this difference into account, we consider the structural model results only. For EU, the expo-power function reduces to a similar CRRA power function in the two laboratory treatments; the relevant risk aversion parameter is not significantly different (risk: $\beta = -0.504$; ambiguity: $\beta = -0.861$; $p = 0.371$). In the limelight, however, the risk aversion parameter of the resulting CARA function is marginally significantly larger in the experiment with ambiguity than in the one without (risk: $\alpha = 0.010$; ambiguity: $\alpha = 0.021$; $p = 0.066$). The CEs (CCs) for a lottery with a 50% chance of €100 illustrate these differences. Under laboratory conditions, the values of €63.07 (126%; risk) and €68.91 (138%; ambiguity) are relatively similar. Under limelight conditions, the values of €38.07 (76%; risk) and €27.72 (55%; ambiguity) are clearly more different.

For PT, the differences in behavior translate into different loss aversion coefficients. For the laboratory treatments, there is no significant difference between the two experiments (risk: $\lambda = 1.259$; ambiguity: $\lambda = 1.505$; $p = 0.349$). For the limelight treatments, however, the coefficient is significantly larger in the one with ambiguity (risk: $\lambda = 1.863$, ambiguity: $\lambda = 2.825$, $p = 0.020$).[11]

IV. Conclusion and Discussion

To analyze how risky choice in the limelight differs from that under more usual experimental laboratory conditions, we conducted two incentivized experiments that mimicked the game of the TV show *Deal or No Deal*. In the laboratory treatments of the experiments, subjects made decisions in a standard, computerized laboratory setting as typically employed in economic experiments. In the limelight treatments, subjects made their choices in a simulated game show environment, which included a live audience, a game show host, and video cameras. The second experiment also had a comeback treatment, in which subjects who had previously gained passive experience by watching others play the game made decisions under laboratory conditions.

We find that subjects are more risk averse in the limelight than in the anonymity of a typical behavioral laboratory. In both experiments, subjects in the limelight have a higher propensity to opt for the sure alternative. For the EU model, this translates into a more concave (risk-averse) utility function. For PT, we observe a higher loss aversion coefficient.

Findings from studies on investor behavior corroborate this result. Barber and Odean (2001, 2002) find that investors trade more and more speculatively after switching from phone-based to online trading. Konana and Balasubramanian (2005) report that investors tend to keep their core investments with traditional brokers and use a small fraction of their wealth to speculate online.

Our second experiment indicates that people in the limelight also have a higher reference point and adjust it more slowly, and that their value function has less curvature. The latter is in line with earlier findings by Miller and Fagley (1991), Takemura (1993, 1994), and Vieider (2011). However, the difference is not significant when we compare subjects in the limelight with experienced subjects in the comeback treatment, suggesting that it may be a spurious effect related to subjects' experience with the game from watching others play. The other results for the comeback treatment reinforce our previous findings about the difference between risk attitudes in and out of the limelight.

While the general degree of risk aversion is affected by the limelight manipulation, the dynamic pattern in risk behavior is not. In particular, and in line with the break-even effect, subjects in and subjects out of the limelight are more risk prone when the game develops substantially worse than expected. Of course, on average, losers faced lower stakes, and increasing relative risk aversion (IRRA) thus appears to be a simple explanation for their greater risk appetite. However, IRRA cannot explain that the choice patterns resemble those found by Post et al. (2008) for games that used stakes of up to 10,000 times the size of those used in our experiment. Furthermore, IRRA would also imply more risk aversion for winners than for subjects in the middle group, which is not what we observe. The risk appetite of winners is in line with the house-money effect: when all possible outcomes are in the gain domain, people no longer feel they might be losing their "own" money and they take more risk.

Our simple PT model allows for a sticky reference point and can capture these path-dependent and very different risk attitudes. All five treatments in our experiments point out that the reference point is sticky and partly determined by subjects' (presumed) initial beliefs about the task outcome. This finding is in line with recent literature on reference-point formation that argues that reference points are expectation based and imperfectly updated (Kőszegi & Rabin, 2006, 2007; Abeler et al., 2011; Baucells et al., 2011; Ericson & Fuster, 2011). For all treatments, the PT model indeed explains subjects' choices significantly better than the EU model that we employ.

The different degree but similar pattern of risk aversion under the two conditions is important in the light of the recent debate on the external validity of laboratory and field

[10] Admittedly, comparisons between the limelight treatments are potentially confounded by another design difference. In contrast to subjects in the other treatments, subjects in the limelight treatment with ambiguity (first experiment) were told at the start that only half of them would play the game. A sense of relief or feelings of luck might thus have influenced the behavior of those selected.

[11] In addition, in the laboratory, the reference point is stickier and more elevated under ambiguity than under risk, and there is marginally significantly less curvature (θ_1: $p = 0.004$; θ_2: $p = 0.009$; α: $p = 0.061$). In contrast, in the limelight, the reference point is less sticky and less elevated under ambiguity than under risk, and there is no significant difference in curvature (θ_1: $p = 0.015$; θ_2: $p = 0.022$; α: $p = 0.707$).

studies (Levitt & List, 2007a, 2007b; Camerer, 2015). Kessler and Vesterlund (2015) argue that while attention has focused on the generalizability of quantitative results, it is much more relevant to focus on the generalizability of qualitative results, as most experimental studies are focused on the direction rather than the magnitude of effects. Furthermore, they argue that while the external validity of quantitative results is highly contested, this is not the case for the external validity of qualitative results. Levitt and List (2007b, p. 351), for example, state that "even for those experiments that are affected by our concerns, it is likely that the qualitative findings of the lab are generalizable, even when the quantitative magnitudes are not." Indeed, a large number of studies suggest that qualitative results generalize between lab and field settings, even if quantitative results differ (Kagel & Roth, 2000; Tenorio & Cason, 2002; Healy & Noussair, 2004; Isaac & Schnier, 2005; Antonovics et al., 2009; Östling et al., 2011; Bolton, Greiner, & Ockenfels, 2013). Our finding of similar patterns of risk taking under different experimental conditions supports the positive view on the generalizability of qualitative results. At the same time, the different degrees of risk taking across conditions sketch a negative picture on the generalizability of quantitative estimates. Where scrutiny has thus far predominantly been considered as a disturbing factor in tasks where moral and wealth are competing objectives (Levitt & List, 2007a, 2007b), this result suggests that scrutiny also affects behavior when moral concerns do not play a role.

The most important difference between the two sets of experiments in our study was that the second used a simple deterministic model for the percentage bank offers that was known to subjects, while subjects in the first set were faced with some uncertainty about the precise offers. Much empirical evidence shows that people are averse to ambiguity or uncertainty about outcome probabilities (Ellsberg, 1961; Camerer & Weber, 1992). When we compare the results of the two experiments, we find that subjects under limelight conditions are indeed more adventurous when the bank offer structure is deterministic and known instead of ambiguous to them, while we find no evidence that behavior under laboratory conditions is affected by this design change. The different effect of ambiguity in and out of the limelight is in line with literature that suggests that ambiguity aversion is related to the presence of outside observers (Curley, Yates, & Abrams, 1986; Trautmann, Vieider, & Wakker, 2008; Muthukrishnan, Wathieu, & Xu, 2009). Also, the absence of an effect of ambiguity under laboratory conditions corresponds with the findings of Fox & Tversky (1995). Through various experiments under conditions of anonymity resembling our laboratory treatments, they find evidence that ambiguity aversion does not occur when there is no contrast of the ambiguous event with a similar but less ambiguous event. Such a contrast is indeed salient in most studies that classify ambiguity aversion as a real phenomenon. In our case, the task did not embed any contrast, and subjects were not aware of any other related experiment with differently generated bank offers.[12,13]

Comparisons between the results of the comeback treatment and the basic laboratory treatment can identify the effect of passive experience on risk tolerance in our experiment. Recent literature shows that experience helps to eliminate anomalous behavior—in particular, loss aversion among market participants (List, 2003, 2004b, 2011; Engelmann & Hollard, 2010; Seru, Shumway, & Stoffman, 2010). We find no clear evidence in this direction, perhaps because the experience of the subjects in the comeback treatment was only passive, because learning is slow and subjects observed only twenty others, or because their choice problems were of a different nature from those in a market context. More specifically, we find that the passive experience that comeback-treatment subjects acquired by watching others play does not affect the loss aversion parameter of our PT model, but we do find evidence for decreased curvature and a more elevated and sticky reference point. Interestingly, although the empirical fit of the PT model is much better than that of the EU model for every treatment, the improvement is strongest for the comeback treatment. This suggests that passive experience strengthens rather than weakens prospect theory like behavior here. A possible explanation is that experience from watching others brings along vivid task-specific expectations and reference, which in turn guide subjects' own behavior. The stickier reference point of experienced subjects indeed points in this direction.

Using DOND as the experimental task has a number of advantages, most notably that the game allows the study of path dependence. However, at the same time, the stop-go nature of DOND might confound the interpretation of our results. In fact, subjects in our limelight treatments were asked to either decide to take risk and stay in the limelight or to opt for a safe money offer and step out of the limelight. As a result, subjects are more likely to "Deal" if they suffer a fixed disutility from being in the limelight. Although we cannot completely rule out this alternative explanation for our results, it does not appear particularly

[12] Interestingly, the comparative ignorance hypothesis of Fox and Tversky (1995) is grounded on the finding of Heath and Tversky (1991) that ambiguity aversion is driven by people's feeling of (in)competence. Possibly, the presence of onlookers in our limelight treatments undermined our subjects' confidence in their capability to perform the task, and this way amplified the effect of the ambiguous bank offer structure on choice.

[13] The uncertainty about the bank offers in our first experiment can be interpreted as a "background risk," although in a strict sense, background risk is mostly regarded and implemented as an additive risk to a subject's overall wealth and not—akin to the uncertainty about future percentage bank offers here—as a multiplicative risk to the outcomes of one choice option only. Based on certain assumptions, most theoretical accounts predict that individuals take less risk in the presence of background risk (Pratt & Zeckhauser, 1987; Gollier & Pratt, 1996; Eeckhoudt, Gollier, & Schlesinger, 1996). Experiments by Harrison, List, and Towe (2007) and Lee (2008) confirm this prediction, whereas the findings of Lusk and Coble (2008) and Herberich and List (2012) indicate that background risk has little to no effect on risky choice.

strong for several reasons. First, self-reflection suggests that such a disutility would rapidly decrease as the game progresses, as many people get used to being in the limelight after a while. When subjects have to make decisions that make a real difference, they have already gone through an introductory talk with the host and played several trivial game rounds. Second, deciding "No deal" commits to playing only one round more, and rounds are short, especially at the critical stages of the game when few or only one case is to be opened. The extra time that would be involved is in fact negligible in the light of the time already spent on stage. Third, and perhaps most important, the data contradict a fixed disutility of being in the limelight. If such a disutility exists, the decisions of the most unfortunate subjects in our sample would be disproportionately affected by it. In our data, we find that losers have a strong tendency to continue to play, and this tendency appears to be even stronger, not weaker, in the limelight than in the laboratory. The latter two arguments similarly indicate that it is unlikely that the treatment effect is driven by a propensity among limelight subjects to reduce the time costs of their fellow students in the audience.

Another potential downside of using DOND is that it may entail a specific demand effect under limelight conditions. As Gertner (1993) pointed out, taking risk is more entertaining for spectators, and this might lead subjects to make riskier choices. In contrast to this intuitive prediction, however, we find that subjects take less risk in the limelight than in the laboratory. Apparently, such a demand effect is relatively unimportant. The overall treatment effect that we find may seem more in line with a demand effect where subjects feel encouraged to make safe choices. However, if going against the perceived demand of spectators generates a fixed disutility, it would again especially affect the behavior of losers, which is not what we observe.

Our experimental design allows us to study the influence of public scrutiny on specific parameters of risky choice models, but it does not allow us to fully disentangle the underlying mechanisms. Next to the demand effects already noted, subjects' fear of others' judgments can be expected to play a role. Indeed, the finding of increased ambiguity aversion under public scrutiny is generally interpreted in terms of justifiability and fear of negative judgments (Curley et al., 1986; Trautman et al. 2008). Similar fears may lead to increased risk aversion. In addition, a number of psychological studies suggest that emotions influence risky choice (Loewenstein et al., 2001; Rick & Loewenstein, 2008). Public scrutiny may entail feelings of stress and anxiety and lead to a state of physiological arousal. Several studies indicate that anxiety lowers subjects' propensity to take risk (Raghunathan & Pham, 1999; Maner et al., 2007; Kuhnen & Knutson, 2011). Similarly, Mano (1994) finds that "distress"—the combination of a negative (unpleasant) emotional state and a high level of arousal—leads to less risk taking. Future research can more directly investigate the mechanisms that underlie the relation between risk tolerance and public scrutiny, as well as the interplay between them.

In sum, our findings provide a mixed message about the generalizability of findings from one setting to another when the degree of public scrutiny is different. Quantitative measurements of risk preferences do not seem to have universal applicability, but the qualitative pattern of path dependence in risk behavior appears to be robust.

REFERENCES

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman, "Reference Points and Effort Provision," *American Economic Review* 101 (2011), 470–492.

Antonovics, Kate, Peter Arcidiacono, and Randall Walsh, "Games and Discrimination: Lessons from *The Weakest Link*," *Journal of Human Resources* 40 (2005), 918–947.

——— "The Effects of Gender Interactions in the Lab and in the Field," THIS REVIEW 91 (2009), 152–162.

Baltussen, Guido, Thierry Post, Martijn J. van den Assem, and Peter P. Wakker, "Random Incentive Systems in a Dynamic Choice Experiment," *Experimental Economics* 15 (2012), 418–443.

Barber, Brad M., and Terrance Odean, "The Internet and the Investor," *Journal of Economic Perspectives* 15 (2001), 41–54.

——— "Online Investors: Do the Slow Die First?" *Review of Financial Studies* 15 (2002) 455–487.

Baucells, Manel, Martin Weber, and Frank Welfens, "Reference-Point Formation and Updating," *Management Science* 57 (2011), 506–519.

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

Belot, Michèle, V. Bhaskar, and Jeroen van de Ven, "Promises and Cooperation: Evidence from a TV Game Show," *Journal of Economic Behavior and Organization* 73 (2010), 396–405.

Bennett, Randall W., and Kent A. Hickman, "Rationality and *The Price Is Right*," *Journal of Economic Behavior and Organization* 21:1 (1993), 99–105.

Berk, Jonathan B., Eric Hughson, and Kirk Vandezande, "The Price Is Right, But Are the Bids? An Investigation of Rational Decision Theory," *American Economic Review* 86 (1996), 954–970.

Blavatskyy, Pavlo, and Ganna Pogrebna, "Models of Stochastic Choice and Decision Theories: Why Both Are Important for Analyzing Decisions," *Journal of Applied Econometrics* 25 (2010), 963–986.

Bolton, Gary, Ben Greiner, and Axel Ockenfels, "Engineering Trust: Reciprocity in the Production of Reputation Information," *Management Science* 59 (2013), 265–285.

Bond, Charles F., and Linda J. Titus, "Social Facilitation: A Meta-Analysis of 241 Studies," *Psychological Bulletin* 94 (1983), 265–292.

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

——— "Does Risk Aversion Vary with Decision-Frame? An Empirical Test Using Recent Game Show Data," *Review of Behavioral Finance* 1:1–2 (2009b), 44–61.

Camerer, Colin F., "The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List," (pp. 249–295), in Guillaume R. Fréchette and Andrew Schotter, eds., *Handbook of Experimental Economic Methodology* (New York: Oxford University Press, 2015).

Camerer, Colin F., and Martin Weber, "Recent Developments in Modeling Preferences: Uncertainty and Ambiguity," *Journal of Risk and Uncertainty* 5 (1992), 325–370.

Curley, Shawn P., J. Frank Yates, and Richard A. Abrams, "Psychological Sources of Ambiguity Avoidance," *Organizational Behavior and Human Decision Processes* 38 (1986), 230–256.

Deck, Cary A., Jungmin Lee, and Javier A. Reyes, "Risk Attitudes in Large Stake Gambles: Evidence from a Game Show," *Applied Economics* 40:1 (2008), 41–52.

Eeckhoudt, Louis, Christian Gollier, and Harris Schlesinger, "Changes in Background Risk and Risk Taking Behavior," *Econometrica* 64 (1996), 683–689.
Ellsberg, Daniel, "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics* 75 (1961), 643–669.
Engelmann, Dirk, and Guillaume Hollard, "Reconsidering the Effect of Market Experience on the 'Endowment Effect,'" *Econometrica* 78 (2010), 2005–2019.
Ericson, Keith M. Marzilli, and Andreas Fuster, "Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments," *Quarterly Journal of Economics* 126 (2011), 1879–1907.
Falk, Armin, and James J. Heckman, "Lab Experiments Are a Major Source of Knowledge in the Social Sciences," *Science* 326:5952 (2009), 535–538.
Fox, Craig E., and Amos Tversky, "Ambiguity Aversion and Comparative Ignorance," *Quarterly Journal of Economics* 110 (1995), 585–603.
Fullenkamp, Connel, Rafael Tenorio, and Robert Battalio, "Assessing Individual Risk Attitudes Using Field Data from Lottery Games," THIS REVIEW 85 (2003), 218–226.
Gertner, Robert, "Game Shows and Economic Behavior: Risk-Taking on *Card Sharks*," *Quarterly Journal of Economics* 108 (1993), 507–521.
Gollier, Christian, and John W. Pratt, "Risk Vulnerability and the Tempering Effect of Background Risk," *Econometrica* 64 (1996), 1109–1123.
Hartley, Roger, Gauthier Lanot, and Ian Walker, "Who Really Wants to Be a Millionaire? Estimates of Risk Aversion from Gameshow Data," *Journal of Applied Econometrics* 29 (2013), 861–869.
Harrison, Glenn W., John A. List, and Charles Towe, "Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion," *Econometrica* 75 (2007), 433–458.
Healy, Paul J., and Charles N. Noussair, "Bidding Behavior in the *Price Is Right* Game: An Experimental Study," *Journal of Economic Behavior and Organization* 54 (2004), 231–247.
Heath, Chip, and Amos Tversky, "Preference and Belief: Ambiguity and Competence in Choice under Uncertainty," *Journal of Risk and Uncertainty* 4:1 (1991), 5–28.
Herberich, David H., and John A. List, "Digging into Background Risk: Experiments with Farmers and Students," *American Journal of Agricultural Economics* 94 (2012), 457–463.
Hersch, Philip L., and Gerald S. McDougall, "Decision Making under Uncertainty When the Stakes Are High: Evidence from a Lottery Game Show," *Southern Economic Journal* 64:1 (1997), 75–84.
Isaac, R. Mark, and Kurt Schnier, "Silent Auctions in the Field and in the Laboratory," *Economic Inquiry* 43 (2005), 715–733.
Kagel, John H., and Alvin E. Roth, "The Dynamics of Reorganization in Matching Markets: A Laboratory Experiment Motivated by a Natural Experiment," *Quarterly Journal of Economics* 115 (2000), 201–235.
Kahneman, Daniel, and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47 (1979), 263–292.
Kameda, Tatsuya, and James H. Davis, "The Function of the Reference Point in Individual and Group Risk Decision Making," *Organizational Behavior and Human Decision Processes* 46 (1990), 55–76.
Kessler, Judd B., and Lise Vesterlund, "The External Validity of Laboratory Experiments: The Misleading Emphasis on Quantitative Effects" (pp. 391–406) in Guillaume R. Fréchette and Andrew Schotter, eds., *Handbook of Experimental Economic Methodology* (New York: Oxford University Press).
Konana, Prabhudev, and Sridhar Balasubramanian, "The Social-Economic-Psychological Model of Technology Adoption and Usage: An Application to Online Investing," *Decision Support Systems* 39 (2005), 505–524.
Kőszegi, Botond, and Matthew Rabin, "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics* 121 (2006), 1133–1165.
——— "Reference-Dependent Risk Attitudes," *American Economic Review* 97 (2007), 1047–1073.
Kuhnen, Camelia M., and Brian Knutson, "The Influence of Affect on Beliefs, Preferences, and Financial Decisions," *Journal of Financial and Quantitative Analysis* 46 (2011), 605–626.
Lee, Jinkwon, "The Effect of the Background Risk in a Simple Chance Improving Decision Model," *Journal of Risk and Uncertainty* 36 (2008), 19–41.
Lerner, Jennifer S., and Philip E. Tetlock, "Accounting for the Effects of Accountability," *Psychological Bulletin* 125 (1999), 255–275.
Levitt, Steven D. "Testing Theories of Discrimination: Evidence from *Weakest Link*," *Journal of Law and Economics* 47 (2004), 431–452.
Levitt, Steven D., and John A. List, "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *Journal of Economic Perspectives* 21 (2007a), 153–174.
——— "Viewpoint: On the Generalizability of Lab Behaviour to the Field," *Canadian Journal of Economics* 40 (2007b), 347–370.
List, John A. "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics* 118 (2003), 41–71.
——— "Young, Selfish and Male: Field Evidence of Social Preferences," *Economic Journal* 114 (2004a), 121–149.
——— "Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace," *Econometrica* 72 (2004b), 615–625.
——— "*Friend or Foe?* A Natural Experiment of the Prisoner's Dilemma," THIS REVIEW 88 (2006), 463–471.
——— "Does Market Experience Eliminate Market Anomalies? The Case of Exogenous Market Experience," *American Economic Review* 101 (2011), 313–317.
List, John A., Sally Sadoff, and Mathis Wagner, "So You Want to Run an Experiment, Now What? Some Simple Rules of Thumb for Optimal Experimental Design," *Experimental Economics* 14 (2011), 439–457.
Loewenstein, George F., "Experimental Economics from the Vantage-Point of Behavioural Economics," *Economic Journal* 109:453 (1999), F25–F34.
Loewenstein, George F., Elke U. Weber, Christopher K. Hsee, and Ned Welch, "Risk as Feelings," *Psychological Bulletin* 127 (2001), 267–286.
Lusk, Jayson L., and Keith H. Coble, "Risk Aversion in the Presence of Background Risk: Evidence from an Economic Experiment" (pp. 315–340), in James C. Cox and Glenn W. Harrison, eds., *Risk Aversion in Experiments* (Somerville, MA: Emerald Group Publishing, 2008).
Maner, Jon K., J. Anthony Richey, Kiara Cromer, Mike Mallott, Carl W. Lejuez, Thomas E. Joiner, and Norman B. Schmidt, "Dispositional Anxiety and Risk-Avoidant Decision-Making," *Personality and Individual Differences* 42 (2007), 665–675.
Mano, Haim, "Risk-Taking, Framing Effects, and Affect," *Organizational Behavior and Human Decision Processes* 57 (1994), 38–58.
Metrick, Andrew, "A Natural Experiment in *Jeopardy!*" *American Economic Review* 85 (1995), 240–253.
Miller, Paul M., and Nancy S. Fagley, "The Effects of Framing, Problem Variations, and Providing Rationale on Choice," *Personality and Social Psychology Bulletin* 17 (1991), 517–522.
Muthukrishnan, A. V., Luc Wathieu, and Alison Jing Xu, "Ambiguity Aversion and the Preference for Established Brands," *Management Science* 55 (2009), 1933–1941.
Oberholzer-Gee, Felix, Joel Waldfogel, and Matthew W. White, "*Friend or Foe?* Cooperation and Learning in High-Stakes Games," THIS REVIEW 92 (2010), 179–187.
Östling, Robert, Joseph Tao-yi Wang, Eileen Y. Chou, and Colin F. Camerer, "Testing Game Theory in the Field: Swedish LUPI Lottery Games," *American Economic Journal: Microeconomics* 3:3 (2011), 1–33.
Post, Thierry, Martijn J. van den Assem, Guido Baltussen, and Richard H. Thaler, "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show," *American Economic Review* 98 (2008), 38–71.
Pratt, John W., and Richard J. Zeckhauser, "Proper Risk Aversion," *Econometrica* 55 (1987), 143–154.
Raghunathan, Rajagopal, and Michel Tuan Pham, "All Negative Moods Are Not Equal: Motivational Influences of Anxiety and Sadness on Decision Making," *Organizational Behavior and Human Decision Processes* 79 (1999), 56–77.
Rick, Scott, and George Loewenstein, "The Role of Emotion in Economic Behavior," (pp. 138–156), in Michael Lewis, Jeannette M. Haviland-Jones, Lisa Feldman Barrett, eds., *Handbook of Emotions*, 3rd ed. (New York: Guilford Press, 2008.
Seru, Amit, Tyler Shumway, and Noah Stoffman, "Learning by Trading," *Review of Financial Studies* 23 (2010), 705–739.

Takemura, Kazuhisa, "The Effect of Decision Frame and Decision Justification on Risky Choice," *Japanese Psychological Research* 35 (1993), 36–40.

——— "Influence of Elaboration on the Framing of Decision," *Journal of Psychology* 128 (1994), 33–39.

Tenorio, Rafael, and Timothy N. Cason, "To Spin or Not to Spin? Natural and Laboratory Experiments from *The Price Is Right*," *Economic Journal* 112:476 (2002), 170–195.

Thaler, Richard H., and Eric J. Johnson, "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice," *Management Science* 36 (1990), 643–660.

Trautmann, Stefan T., Ferdinand M. Vieider, and Peter P. Wakker, "Causes of Ambiguity Aversion: Known versus Unknown Preferences," *Journal of Risk and Uncertainty* 36 (2008), 225–243.

Tversky, Amos, and Daniel Kahneman, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5 (1992), 297–323.

van den Assem, Martijn J., Dennie van Dolder, and Richard H. Thaler, "Split or Steal? Cooperative Behavior When the Stakes Are Large," *Management Science* 58 (2012), 2–20.

van Dolder, Dennie, Martijn J. van den Assem, Colin F. Camerer, and Richard H. Thaler, "Standing United or Falling Divided? High Stakes Bargaining in a TV Game Show," *American Economic Review* 105 (2015), 402–407. http://dx.doi.org/10.1257/aer.p20151017.

Vieider, Ferdinand M, "The Effect of Accountability on Loss Aversion," *Acta Psychologica* 132:1 (2009), 96–101.

——— "Separating Real Incentives and Accountability," *Experimental Economics* 14 (2011), 507–518.

Weigold, Michael F., and Barry R. Schlenker, "Accountability and Risk Taking," *Personality and Social Psychology Bulletin* 17:1 (1991), 25–29.

Wooldridge, Jeffrey M., "Cluster-Sample Methods in Applied Econometrics," *American Economic Review* 93 (2003), 133–138.

Zajonc, Robert B., "Social Facilitation," *Science* 149:3681 (1965), 269–274.