

# Expectations, Satisfaction, and Utility from Experience Goods: A Field Study in Theaters\*§

PEDRO REY-BIEL

ESADE Business School  
Universitat Ramon Llull

AYELET GNEEZY

Rady School of Business,  
University of California-San Diego

URY GNEEZY

Rady School of Business,  
University of California-San Diego

DOMINIQUE LAUGA

Judge Business School,  
University of Cambridge

JOAN LLULL

MOVE,

Universitat Autònoma de Barcelona,  
and Barcelona GSE

February 2020

*Understanding what affects consumption satisfaction is fundamental to understanding consumer behavior. Measuring satisfaction, however, is not trivial, especially in the context of experience goods where perceived quality is often subjective and unobservable prior to consumption. We report the results of a field study ( $N = 433$ ) conducted in collaboration with a theater that uses pay-what-you-want pricing, inviting the audience to pay at the end of the show. We find that survey measures of post-consumption enjoyment capture only part of consumer satisfaction. This occurs because individuals with a larger positive gap between self-reported expected and actual enjoyment pay significantly more. Our analyses further indicate that neither expected nor realized enjoyment predict payments once we control for the expectation-realization gap. The paper highlights the managerial implications of these findings.*

**Keywords:** *Measuring satisfaction; experience goods; field experiment; pay-what-you-want; expectations*

---

\* Corresponding author: Pedro Rey-Biel, ESADE Business School, Universitat Ramon Llull. Av. de la Torre Blanca, 59, Sant Cugat E-08172 Barcelona, Spain. Tel: (+34) 932 806 162 (Ext. 5560). E-mail: pedro.rey [at] esade.edu.

§ We are grateful to Sala Beckett and Sixto Paz Produccions for their collaboration on this project. We thank Miguel A. Ballester, Marco Bertini, Guillaume Hollard, Steffen Huck, Alex Imas, Wieland Müller, Marta Serra-García, Sigrid Suetens, Matthias Sutter and seminar participants at several institutions for their comments, and Adrià Bronchal, Elena Costas-Pérez, Sergi Martínez, Dilan Okçuoğlu, and Javier Rodríguez-Camacho for excellent research assistance. This work has been supported by the Ministerio de Economía y Competitividad through the Ramón y Cajal program, and grants ECO2012-31962, ECO2014-59056-JIN, ECO2015-63679-P, and the Severo Ochoa Programme for Centres of Excellence in R&D, SEV-2015-0563; the Fundación Ramón Areces; and the Generalitat de Catalunya (2017-SGR-1765).

## I. Introduction

Consumption satisfaction represents one of the fundamental aspects of consumer behavior, and understanding its determinants could help us shine a light on an important part of economic activity. Conceptually, satisfaction can be thought of as experienced utility –the hedonic experience of an outcome (e.g. Kahneman, Wakker and Sarin, 1997; Kahneman and Thaler, 2006). Measuring it is far from trivial, especially when the perceived quality is unobservable prior to consumption (Nelson, 1970), and sometimes also post consumption (Balafoutas, Beck, Kerschbamer and Sutter, 2013; Darby and Karni, 1973; Dulleck, Kerschbamer and Sutter, 2011).<sup>1</sup>

Research suggests that consumers’ expectations prior to consumption, and the extent to which consumption confirms these expectations, play a major role in their post-consumption satisfaction (Schwartz, 2005; Köszegi and Rabin, 2006; Gneezy et al., 2014). Yet, traditional measures of experienced utility, such as consumers’ self-reported enjoyment, do not capture the interaction of expectations and satisfaction. This literature provides indirect evidence for reference-dependence in utility functions, based on how the intrinsic value of an outcome compares to expectations (Kahneman and Tversky, 1979; Medvec, Madey and Gilovich, 1995; Card and Dahl, 2011; Abeler, Falk, Goette and Huffman, 2011; Bushong and Gagnon-Bartasch, 2017).

We develop and test a novel approach that considers three factors contributing to the difficulty of measuring experienced utility. First, participants must consume the product whose consumption satisfaction the researcher wants to impute, preferably in a naturally occurring environment. Second, the researcher needs individual-level data to measure each participant’s expectations prior to consumption, as well as satisfaction post-consumption. Finally, the researcher must use a clean and reliable measure of satisfaction post-consumption. We address these challenges using a unique field study conducted in a theater setting. Participants were theatergoers who completed a brief survey before and after the show, providing us with individual level data. Our sample consisted of individuals consuming a good of their choice in a natural environment. The theater uses

---

<sup>1</sup> Throughout the paper, we refer to three similar even though not exactly equivalent concepts: satisfaction, enjoyment, and utility. Consumer satisfaction refers to the fulfillment of one’s wishes, expectations, or needs, or the pleasure derived from this fulfillment. Enjoyment and utility are similar concepts, but they are also defined for hypothetical situations, and, therefore, they are not necessarily associated with actual fulfillment. Even though they are harder to distinguish from each other conceptually, in this paper we use the term enjoyment to refer to our empirical measurement of utility, expected or realized, based on our surveys, whereas we use the term utility in our theoretical derivations.

a “pay-what-you-want” (PWYW) pricing scheme (Gneezy, Gneezy, Nelson and Brown, 2010), leaving the audience free to choose whether and how much to pay at the end of the play. We use these payments as our measure of experienced utility, and link it to the questionnaire data (see Lynn, 2006; Azar, 2007).

Our analyses show that survey measures of post-consumption satisfaction capture only part of the picture. What matters most is the difference between these measures and individuals’ expectations before seeing the play: individuals with a greater gap between self-reported measures expected and actual enjoyment pay significantly more. Once the enjoyment gap is accounted for, neither individuals’ expectations nor self-reported enjoyment predict payments more accurately.

From a practical standpoint, these findings have important implications for how experience goods are marketed and priced. In particular, by demonstrating the critical role of expectations, our findings highlight an important trade-off marketers should consider: appealing to a larger set of consumers by increasing expectations runs the risk of disappointing them. From a methodological point of view, PWYW pricing provides a good measure of the perceived quality of the experience and may positively affect revenue compared to traditional pricing.

The rest of the paper is organized as follows. In the next section, we present a model where consumers have reference-dependent utility and we analyze how their experienced utility varies with realized enjoyment, expected enjoyment and the enjoyment gap. Section III then presents the field experiment run in a theater and the results are analyzed in Section IV. Section V summarizes the key findings and offers managerial implications.

## II. Theoretical Model

When evaluating the consumption of an experience good, consumers consider the outcome of the experience itself as well as how that outcome compares with their reference point formed by their expectations. The two elements are captured by the model developed by Köszegi and Rabin (2006). Specifically, the reference-dependent utility  $u(c|r)$  of consuming good  $c$  with a reference point  $r$  is the sum of the consumption utility  $m(c)$  and the gain-loss utility  $\mu(m(c) - m(r))$ . The gain-loss utility is based on the gap  $m(c) - m(r)$  between the realized consumption utility  $m(c)$  and the expected consumption utility  $m(r)$ . A positive gap means that the realized outcome exceeds the expected one and the consumer derives some extra utility; a negative gap creates dis-utility, because the realized outcome falls short of expectations. We assume that  $\mu$  satisfies the usual properties of a value function (Kahneman and Tversky, 1979). We assume that the gain-loss

utility function is  $\mu(x) = \eta x$  for  $x \geq 0$  and  $\mu(x) = \eta\lambda x$  for  $x \leq 0$ , where  $\eta \geq 0$  is the weight associated to the gain-loss utility function, and  $\lambda > 1$  is the coefficient of loss aversion.

We consider a consumer who is uncertain about the quality of an experience good, which can either be high quality  $c_H$  with probability  $\alpha$  or low quality  $c_L$  with probability  $1 - \alpha$ , where  $\alpha \in [0, 1]$ . This distribution forms the reference point  $r$  in her reference-dependent utility. The overall utility of the good of quality  $c_i$  (for  $i \in \{H, L\}$ ) given the reference point  $r$  is given by:

$$u(c_i|r) = m_i + \alpha\mu(m_i - m_H) + (1 - \alpha)\mu(m_i - m_L), \quad (1)$$

where  $m_L \equiv m(c_L)$  and  $m_H \equiv m(c_H)$  denote the two levels of consumption utility. By definition of low/high quality, we have  $m_L < m_H$ . Therefore, when the realized consumption utility is  $m_H$ , the overall utility  $u(c_H|r)$  is larger than its counterpart  $u(c_L|r)$  when the realization is  $m_L$ , that is  $u(c_H|r) > u(c_L|r)$ . Indeed a realized utility exceeding the expected utility means that both the consumption and gain-loss utilities are larger than when realized utility falls short of the expected utility. Further, because the gain-loss utility is non-decreasing, the reference-dependent utility increases with respect to the realized consumption utility  $m_i$ :

$$\partial u(c_i|r)/\partial m_i > 0. \quad (2)$$

To analyze the impact of the expected consumption utility  $\mathbb{E}[m] = \alpha m_H + (1 - \alpha)m_L$  on the overall utility, rewrite the latter as a function of the realized consumption utility  $m_i$ , and the expected consumption utility  $\mathbb{E}[m]$ :

$$\begin{aligned} u(c_H|r) &= m_H + \eta(m_H - \mathbb{E}[m]) \\ u(c_L|r) &= m_L + \eta\lambda(m_L - \mathbb{E}[m]). \end{aligned} \quad (3)$$

From these expectations, we clearly see that the overall utility decreases with expected utility, controlling for realized utility  $m_i$ , that is:

$$\partial u(c_i|r)/\partial \mathbb{E}[m] < 0. \quad (4)$$

We now define the gap between realized and expected utilities as  $\Delta_i \equiv m_i - \mathbb{E}[m]$  and evaluate its impact on the overall utility. First note that  $\Delta_L < 0 < \Delta_H$  capturing that a low consumption utility falls short of expectations while a high consumption utility exceeds expectation. Rewriting the overall utility as a function of  $\Delta_i$  and either realized utility  $m_i$  or expected utility  $\mathbb{E}[m]$  yields:

$$\begin{aligned} u(c_H|r) &= m_H + \eta\Delta_H = \mathbb{E}[m] + (1 + \eta)\Delta_H \\ u(c_L|r) &= m_L + \eta\lambda\Delta_L = \mathbb{E}[m] + (1 + \eta\lambda)\Delta_L. \end{aligned} \quad (5)$$

Keeping the realized consumption utility  $m_i$  or the expected consumption  $\mathbb{E}[m]$  fixed, the reference-dependent utility increases with the expectation-realization gap:

$$\partial u(c_i|r)/\partial \Delta_i > 0. \quad (6)$$

We also find that  $\partial u(c_i|r)/\partial \Delta_i$  is larger for a negative gap  $\Delta_L$  than a positive one  $\Delta_H$ . Specifically, the slope for negative gaps is steeper than the slope for the positive gaps due to loss aversion (since  $\lambda > 1$ ).

Finally, controlling for the gap, the overall utility increases with either the expected or realized utility. Note that this effect becomes negligible when  $\eta$ , the weight associated to the gain-loss utility, becomes very large. Therefore, if the gain-loss utility is the main factor in the overall utility then we find that neither expected nor realized consumption utilities predict overall utility once we control for the expectation-realization gap.

The remainder of the paper reports analyses of data obtained in a field study in which we measured theater goers' expected enjoyment levels ( $\mathbb{E}[m]$ ), realized enjoyment levels ( $m_i$ ), and observed their payments under a PWYW pricing scheme (proxy for overall utility  $u(c_i|r)$ ). Using our theoretical model, our analyses consider five hypotheses derived from the theoretical model. The first three hypotheses link payments and different measures of enjoyment (expected, realized, and the enjoyment gap). First, high expectations might lower the overall enjoyment (see Equation (4)):

**Hypothesis 1** *Payments decrease with expected enjoyment.*

Second, actual reported enjoyment may affect payment (see Equation (2)):

**Hypothesis 2** *Payments increase with realized enjoyment.*

Third, the gap (realized enjoyment minus expected enjoyment) might predict payment (see Equation (6)):

**Hypothesis 3** *Payments increase with the enjoyment gap.*

The remaining three hypotheses shed light on the factors driving the value of an experience good:

**Hypothesis 4**

- (a) *Controlling for enjoyment gap, payments increase with realized enjoyment.*
- (b) *Controlling for enjoyment gap, payments are independent of realized enjoyment when the gain-loss utility is the main factor in the overall utility.*

Finally, consistent with prior research, we predict an asymmetry in the effect of positive versus negative enjoyment gaps.

**Hypothesis 5** *Payments exhibit loss aversion with a slope for negative enjoyment gaps steeper than the slope for positive enjoyment gaps.*

Our results confirm that while both expected enjoyment and realized enjoyment are important drivers of payments on their own (H1 and H2), what really matters is the difference between them—the enjoyment gap (H3). In particular, we find that neither expected nor realized enjoyment levels have a significant impact on payment after controlling for enjoyment gap size, (H4b), suggesting the gain-loss utility is the main factor in the overall utility. Finally, we find that when realized enjoyment is lower than expected enjoyment, consumers appear disappointed, since we observe a decrease in WTP that is greater in absolute terms than the increase in WTP for an equivalent positive gap (H5), consistent with loss aversion.

### III. Methods

We collected the data at the Sala Beckett theater in Barcelona during January-March of 2015. The play —The Effect, by Lucy Prebble— had 40 fully booked performances. Our data refers to the last 19 shows. The producers, Sixto Paz Productions, had been successfully using PWYW pricing for all plays since 2011. Traditionally, audience members pre-book tickets at no cost, knowing that they will be asked to pay any amount they see fit, including zero, upon leaving the theater. We maintained this payment system throughout the study.

A total of 1,962 individuals attended the 19 performances during which we collected data for the study. This data collection process was part of a field experiment. To achieve randomization, a member of our research team approached individuals arriving to the theater in regular time intervals (2 minutes) and asked them to participate in a study. Each night, we invited approximately one-third of the audience to complete two short questionnaires: one before and the other after the show. We used these questionnaires to measure expectations, enjoyment, and demographic variables.

Our request seemed natural because the producers regularly ask the audience to take part in questionnaires, games, or small focus groups before and after their plays. For this show, the producers wanted each participant to randomly pick one of two differently colored pills from a bowl. Once they made their choice, they were told it was either a pill “helping us to study whether it made them enjoy the play more” or a placebo. We find no statistically significant effect of

taking the pill on the distribution of answers to any questionnaire-based measure or on payments. See Appendix for Figure A1 for the questionnaires that were administered, both in original language (Catalan) and an English translation.

We informed participants that the show was being video recorded, and that we needed them to sign a consent form agreeing to be seated in an area where the camera could capture their image. We had originally planned to analyze the relationship between facial expressions, satisfaction, and payments. Unfortunately, the high level of noise in the data prevents us from using it.

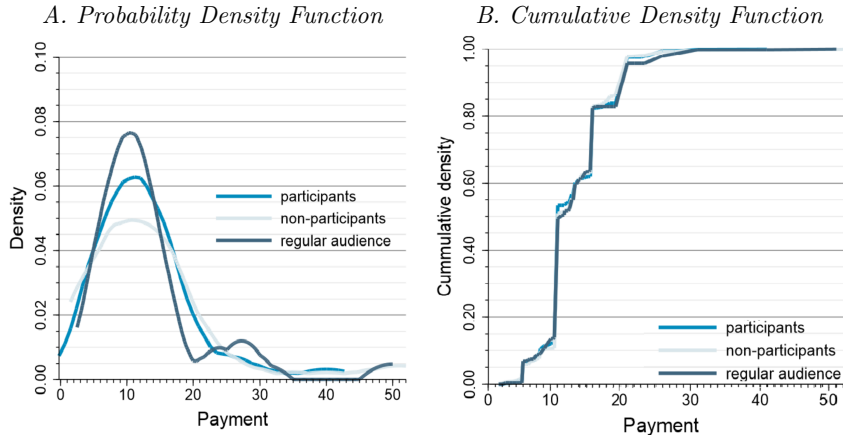
Of the 683 individuals approached, 629 (92%) agreed to participate and were handed the pre-performance questionnaire, which also assigned them a seat. We asked participants to complete the survey before the performance started, and to keep it with them until the end of the show. In addition to basic demographics (age, gender, occupation), participants indicated (using a 7-point scale) the extent to which they expected to enjoy the show. Participants also reported how often they go to plays, whether they or someone else purchased their ticket, and how many individuals were included in their reservation.

At the end of the play, all visitors placed their payments into one of two boxes located on a table inside the stage hall. The only difference between the study participants and other audience members was that the former placed their payment on top of their questionnaire, allowing us to trace it back to their seat numbers.

Once they exited the stage hall, study participants received a second questionnaire asking them to indicate the extent to which they enjoyed watching the play (using the same 7-point scale used in the first questionnaire). The questionnaire also asked the participants how likely they were to recommend the show to others, whether their expectations were met, and the extent to which they agreed with the message of the play –the use of drugs to treat psychiatric conditions. Finally, participants could provide additional comments (open-ended). After completing this second questionnaire, participants left the theater.

Some participants ( $N = 106$ ; 14.3%) failed to complete at least one of the questionnaires, misplaced their payments or paid 0 euros; this figure is similar to the proportion of non-payers among non-participants on all 40 shows. The remaining analyses use the reduced sample of participants for whom we can link payment and questionnaire data at the individual level ( $N = 433$ ).

FIGURE 1. PAYMENTS DISTRIBUTION (Participants, Non-Participants, and Regular Audience)



*Note:* The figure plots the distribution of payments for participants (blue), non-participants (light blue), and regular audience members (dark blue). The left panel shows kernel density estimates; the right panel depicts the empirical cumulative distribution functions for the three samples. A Kolmogorov-Smirnov test of the hypothesis that pairs of the empirical distributions come from the same population cannot reject the null hypothesis. P-values are 0.60, 0.42, and 0.75 for the comparison of participants and non-participants, participants and regular audience, and non-participants and regular audience, respectively.

## IV. Results

### A. Participants, non-participants, and regular nights' audience

First, we tested whether participation in our study affected payments. Of the 3,907 spectators attending the 40 performances of the play, 3,418 (87.5%) made a positive payment. Figure 1 compares payments from our study’s participants who made a positive payment ( $N = 539$ ; 85% of participants) with positive payments from non-participants during the 19 nights of our study ( $N = 1,221$ ), and with positive payments from audiences that visited the theater in the preceding 21 nights ( $N = 1,658$ ; labeled “regular audience”). Figure 1 shows the probability density and cumulative distribution functions of payments for all three groups. The empirical probability distribution function of payments for the three groups is nearly identical, implying that our study did not influence payments. Kolmogorov-Smirnov tests of the hypothesis that any of the possible pairs of empirical distributions come from the same population distribution do not reject the null hypothesis in any case.

During the last four nights of the play, we were allowed to distribute our questionnaires to some non-participants, providing us with additional 148 observations for the pre-performance questionnaire and 197 observations for the post-performance one. We could not, however, link these two questionnaires or the payment information to individual audience members. Our analyses show that



TABLE 1—DESCRIPTIVE STATISTICS

	Average	Std. Dev.	Min	Max
Age	37.45	13.03	16.0	88.0
Female	0.64	0.48	0.0	1.0
Paid in group	0.62	0.49	0.0	1.0
Number of accomp. persons	2.39	1.48	0.0	5.0
First time at venue	0.46	0.50	0.0	1.0
Times at venue before	2.15	3.11	0.0	10.0
First time theatre in a year	0.09	0.28	0.0	1.0
Times in theatre last year	8.11	12.68	0.0	90.0

*Note:* The table reports means, standard deviations, and extremum values for a set of demographic characteristics and theater-attendance habits of participants.

there were no differences, on any measure, between study participants and other audience, with two exceptions: study participants were older (4.4 years on average, with a standard error of 0.84) and attended the show in larger groups (0.43 additional group members, with a standard error of 0.14; see Appendix, Figures A2 through A4). We control for these variables in all analyses.

### B. Descriptive statistics

Table 1 presents the descriptive statistics of demographic variables and theater-attendance habits. Participants were between 16 to 88 years old (age = 37; 64% female). Approximately half the participants (54%) had previously attended other shows at the same venue. Most participants (91%) had attended several shows during the preceding year. In cases where payments were made in groups (62%; average group size = 3.39), we assign an equal fraction of the total payment to each group member. Table 2 shows the descriptive statistics of expectations, enjoyment, and payment data. Payments range from 2.50 to 35 euros, with an average of 12.86 euros, which is above the typical price of 10 euros for independent theater plays in Barcelona.

Participants' reported expected and realized enjoyment show that, in general, attendants expected a high-quality show and enjoyed the experience. Expected enjoyment does not significantly correlate with the variables reported in Table 1, with two exceptions. First, female participants indicated significantly higher expected and realized enjoyment than male participants, although gender is not significantly correlated with the enjoyment gap. Second, expected enjoyment is correlated with whether participants were visiting the venue for the first time. Participants booking tickets themselves did not show significantly higher average expectations than those who did not.

In response to the direct question on whether the show met their expecta-

TABLE 2—DESCRIPTIVE STATISTICS: PAYMENT, EXPECTATIONS, AND ENJOYMENT

	Average	Std. Dev.	Min	Max
Payment	12.86	4.96	2.5	35.0
Reported enjoyment measures:				
Expected enjoyment	6.20	0.81	3.0	7.0
Effective enjoyment	5.92	0.99	2.0	7.0
Other reported measures:				
Expectations were met	3.04	0.59	1.0	4.0
Likelihood of recommend.	3.55	0.68	0.0	4.0
Agreement with message	5.73	1.09	0.0	7.0

*Note:* The table reports means, standard deviations, and the range of values (min and max) for the different variables used in the analysis. Except for "other reported measures", statistics are computed for the sub-sample of 433 participants for which information for all variables is available. The sub-samples used to compute each of the three last rows includes, respectively, 433, 434, and 419 observations.

TABLE 3—CONTINGENCY TABLE OF EXPECTED AND REALIZED ENJOYMENT

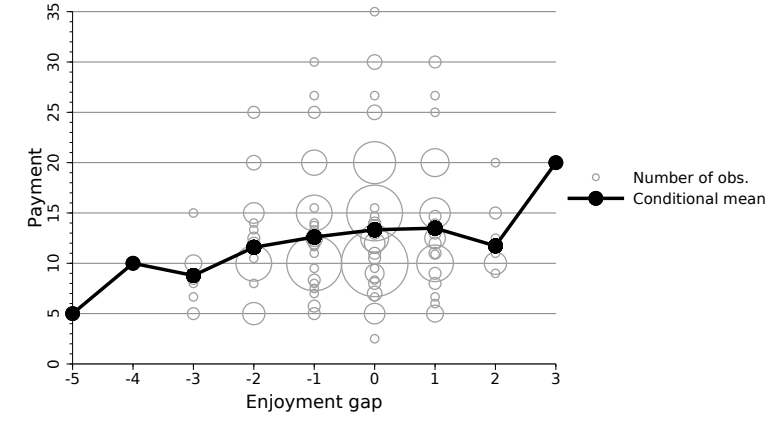
	Expected enjoyment						Total	Below expect.	Above expect.
	2	3	4	5	6	7			
2	0	0	0	2	0	1	3	3	0
3	0	0	0	2	2	1	5	5	0
4	0	0	1	9	11	7	28	27	0
5	0	0	0	27	27	31	85	58	0
6	0	1	1	38	76	61	177	61	40
7	0	0	1	12	36	86	135	0	49
Total	0	1	3	90	152	187	433		
Below expect.:	0	0	0	13	40	101		154	
Above expect.:	0	1	2	50	36	0			89

*Note:* Cell indicate the number of observations that indicated the corresponding levels of expected and realized enjoyment. The row and column labeled "below expectations" represent the number of individuals whose expected enjoyment was lower than their realized enjoyment; the row and column labeled "above expectations" represent the number of individuals whose expected enjoyment was higher than their realized enjoyment.

tions, the participants indicated a mean score of 3.04 (on a 4-point scale), and the likelihood of recommending the show to others averaged 3.55 (on a 4-point scale). Finally, the degree of agreement with the play's message averaged 5.73 (on a 7-point scale). Different measures might capture similar aspects of satisfaction. The highest correlation coefficient among our measures is between realized enjoyment and likelihood of recommendation (0.69, standard error 0.04).

Table 3 summarizes the frequency of all combinations of expected and realized enjoyment observed in the data. As can be seen, most observations correspond to high values of expectations and realized enjoyment. Of the 433 participants, 190 (44%) lie on the diagonal, meaning the show met their expectations. The show exceeded the expectations of 89 (21%) participants and fell short of the

FIGURE 2. DISTRIBUTION OF PAYMENTS CONDITIONAL ON ENJOYMENT GAP



*Note:* The figure shows the distribution of payments conditional on the enjoyment gap. The position of each bubble represent an amount paid at a corresponding enjoyment gap level. The size of the bubble indicates the number of individuals in a given pair. The black line plots the average payment conditional on the enjoyment gap.

expectations of 154 (36%) participants. Enjoyment gap –realized enjoyment minus expected enjoyment– was rarely larger than 2 points (on our 7-point scale).

### C. Payments conditional on the enjoyment gap

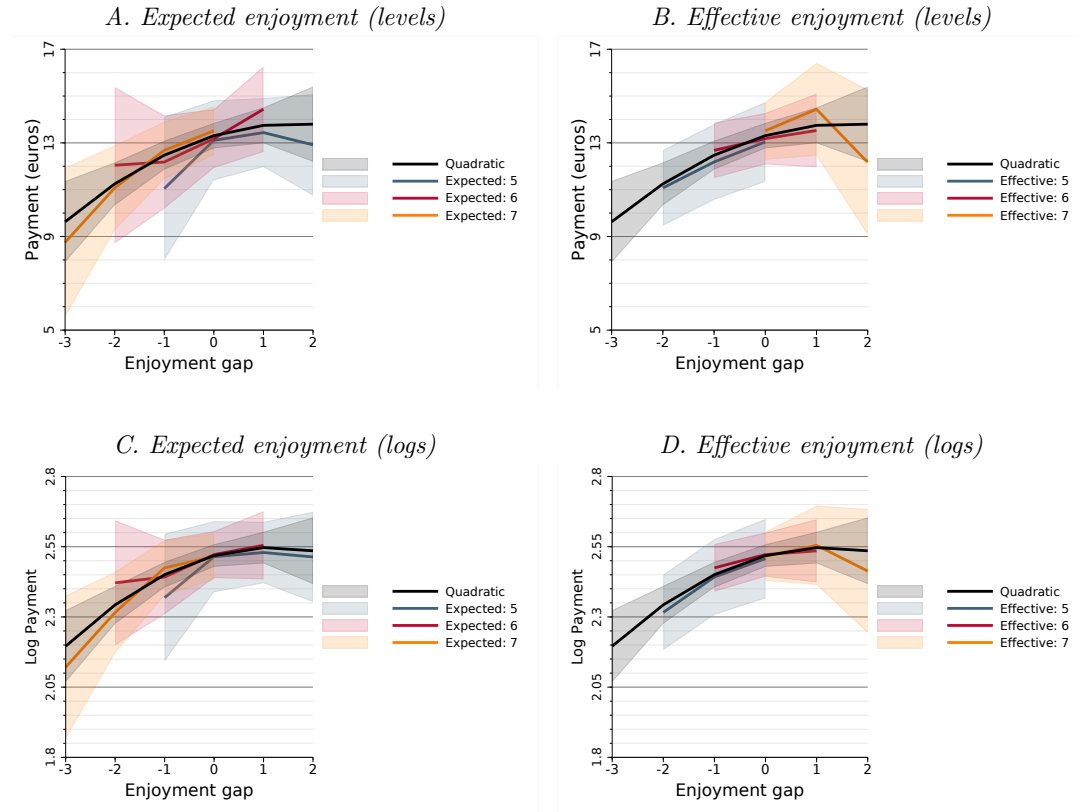
Figure 2 shows the distribution of payments conditional on the enjoyment gap. This distribution has an upward trend. This finding supports Hypothesis 3 that payments increase as the enjoyment gap increases.

Figure 3 further corroborates this finding. The four panels depict the average payment (in levels in the upper panels, in logarithms in the lower panels) as the enjoyment gap increases, conditional on the different levels of expected enjoyment (left panels) or realized enjoyment (right panels). The black solid lines depict the fitted values of a quadratic regression of payment (or log payment) on the enjoyment gap, showing an increasing trend, which is confirmed by the 95% confidence intervals (shown in dotted lines). Figure 3 renders support to our main prediction—the fact that the increasing colored lines overlap, alongside the results of our regression (reported below), confirms payments are determined primarily by the enjoyment gap, over and above self-reported realized enjoyment.

### D. Regression results

The results of a regression of log payment on our measures of enjoyment and the controls summarized in Table 1 are shown in Table 4. All of the reported results are also robust to a specification using amount paid as the dependent variable (reported in the Appendix). We use a logarithmic specification because this is less sensitive to extreme observations. Columns 1 through 4 of Table 4 do not

FIGURE 3. AVERAGE PAYMENT (LEVEL AND LOG) BY ENJOYMENT GAP AND LEVEL



*Note:* Colored solid lines in each plot represent average payment (top plots) or log payment (bottom plots) for each enjoyment-gap level, for the sub-samples of individuals with the indicated expected or realized enjoyment. Only averages of cells with more than five observations are reported in this figure. Black solid lines depict the fitted values of a regression of payment (or log payment) on a second-order polynomial on the enjoyment gap. Shaded areas indicate 95% confidence intervals. The figure was derived based on the regression results reported below.

include night fixed effects, whereas Columns 5 and 6 do.

Among the controls, the age coefficient is always positive and significant, indicating older people pay more, perhaps because age may serve as a proxy for income. Column (1) shows expected enjoyment alone is not a significant determinant of payment (rejecting H1). Column 2 shows realized enjoyment is a significant determinant of payment (consistent with H2). Including both regressors (Column 3) improves the fit (the adjusted  $R^2$  increases from 0.149 in Column 2 and 0.125 in Column 1 to 0.162) and both coefficients become significant. Interestingly, the two coefficients are of opposite sign and statistically indistinguishable in magnitude ( $p = 0.479$ ), suggesting that lowering expectations by one unit has the same effect on payment as increasing realized enjoyment by one unit. This finding is in line with the results presented in Figure 3, where the payment curves by expected and realized enjoyment levels lie on top of each other, and is confirmed by the significant and positive coefficient of the enjoyment gap (Column 4), show-

TABLE 4—REGRESSION RESULTS: LOG PAYMENT ON ENJOYMENT LEVELS AND GAPS

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	2.407 (0.177)	1.834 (0.139)	2.125 (0.188)	2.240 (0.075)	1.946 (0.214)	2.116 (0.106)
Enjoyment gap	—	—	—	0.071 (0.018)	—	0.069 (0.018)
Expected enjoyment	-0.035 (0.025)	—	-0.059 (0.025)	—	-0.051 (0.027)	—
Realized enjoyment	—	0.067 (0.020)	0.078 (0.020)	—	0.079 (0.020)	—
Age	0.010 (0.002)	0.010 (0.002)	0.010 (0.002)	0.010 (0.002)	0.010 (0.002)	0.010 (0.002)
Female	0.012 (0.032)	-0.024 (0.032)	-0.009 (0.033)	-0.002 (0.030)	-0.013 (0.034)	-0.003 (0.031)
Paid in group	-0.063 (0.041)	-0.075 (0.040)	-0.069 (0.040)	-0.067 (0.040)	-0.053 (0.040)	-0.050 (0.040)
Num. of accomp. pers. (base=1):						
0	0.069 (0.167)	0.043 (0.160)	0.046 (0.167)	0.049 (0.169)	0.080 (0.168)	0.083 (0.170)
2	-0.102 (0.078)	-0.087 (0.077)	-0.101 (0.074)	-0.105 (0.074)	-0.091 (0.077)	-0.097 (0.076)
3	-0.061 (0.056)	-0.049 (0.055)	-0.053 (0.054)	-0.055 (0.055)	-0.027 (0.060)	-0.031 (0.060)
4	-0.081 (0.055)	-0.085 (0.056)	-0.090 (0.056)	-0.090 (0.056)	-0.086 (0.057)	-0.088 (0.057)
5 or more	-0.069 (0.065)	-0.057 (0.061)	-0.060 (0.061)	-0.062 (0.062)	-0.067 (0.067)	-0.070 (0.068)
First time at venue	-0.038 (0.040)	-0.043 (0.040)	-0.056 (0.039)	-0.057 (0.039)	-0.024 (0.041)	-0.026 (0.040)
First time at theater this year	-0.057 (0.073)	-0.080 (0.071)	-0.083 (0.070)	-0.081 (0.071)	-0.108 (0.069)	-0.103 (0.071)
Night fixed effects	No	No	No	No	Yes	Yes
Expected+Realized=zero (p-val)	—	—	0.479	—	0.319	—
Restricted vs unrestricted (p-val)	—	—	—	0.497	—	0.315
Adjusted R-squared	0.125	0.149	0.162	0.162	0.188	0.187
Num. of observations	433	433	433	433	433	433

*Note:* The table includes regression coefficients for a set of regressions of log payment on a set of controls, reported enjoyment (expected and/or realized), and/or the enjoyment gap, as indicated. Night fixed effects are included in Columns 5 and 6, as indicated. Reported p-values correspond to tests of the null hypothesis that the coefficients of expected and realized enjoyment are equal in absolute value and of opposite sign, and to a test comparing the restricted and unrestricted models. Standard errors clustered by joint-payment groups are reported in parentheses.

ing a 7% increase in payment (1 euro on average) for each additional positive enjoyment-gap unit.

We cannot reject the hypothesis that the unrestricted and restricted models (Columns 3 and 4, respectively) are identical ( $p = 0.497$ ). Including fixed effects in Columns 5 and 6 increases the  $R^2$ , but the estimated coefficients remain virtually unchanged. This finding confirms that the enjoyment gap is a sufficient statistic of expected and realized enjoyment in the payment function.

TABLE 5—REGRESSION RESULTS: NON-LINEAR SPECIFICATIONS

	(1)	(2)	(3)	(4)	(5)	(6)
Enjoyment gap	0.059 (0.025)	0.051 (0.027)	0.050 (0.027)	0.046 (0.028)	0.042 (0.028)	0.036 (0.030)
Enjoyment gap squared	—	—	-0.016 (0.008)	-0.017 (0.008)	-0.011 (0.012)	-0.010 (0.011)
Enjoyment gap cubed	—	—	—	—	0.002 (0.003)	0.002 (0.003)
Realized enjoyment	0.020 (0.028)	0.029 (0.029)	0.005 (0.028)	0.013 (0.029)	0.007 (0.028)	0.018 (0.029)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Night fixed effects	No	No	No	No	Yes	Yes
Slope for gap is zero (p-val)	0.022	0.059	0.002	0.007	0.002	0.020
Level coeff. is zero (p-val)	0.479	0.319	0.846	0.656	0.799	0.545
Adjusted R-squared	0.162	0.188	0.166	0.095	0.165	0.190
Num. of observations	433	433	433	433	433	433

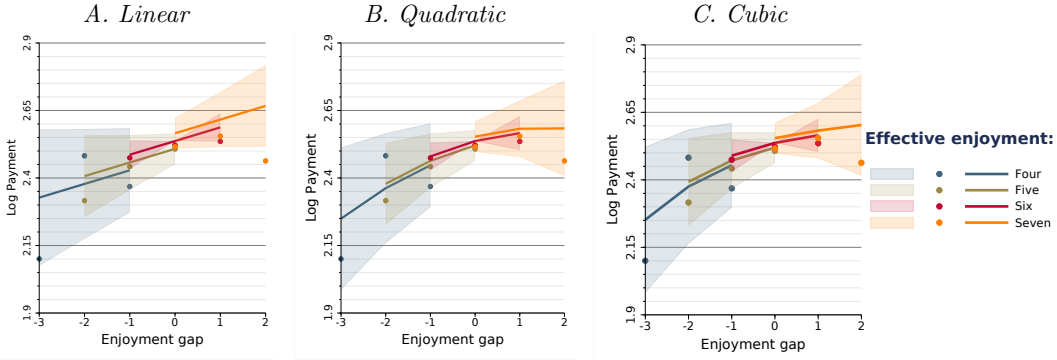
*Note:* The table reports gap polynomials and level coefficients from regressions of log payment on these variables, a set of controls, and, whenever indicated, night fixed effects. Included controls coincide with those in Table 4. Reported p-values correspond to tests of joint significance of the gap polynomial coefficients, and of individual significance of the coefficient of realized enjoyment. Standard errors, clustered by joint-payment groups, are reported in parentheses.

Figure 3 suggests that the relation between enjoyment gaps and payments is likely non-linear. Table 5 studies the nature of this relationship, introducing quadratic and cubic polynomials in the enjoyment gap and the coefficients on the levels in the regressions. Estimated coefficients provide evidence supporting a quadratic relationship but not a cubic one. This non-linearity resembles the classic loss-aversion result (Thaler, 1980), where losses weigh more than equivalent gains. A test of joint significance of the gap polynomial coefficients rejects that the slope is zero at the 5% level. The result is significant at the 10% level for the linear model including night fixed effects. We cannot reject that the estimated level coefficients of realized enjoyment are zero, confirming that conditional on expected enjoyment, estimated payment is the same for the same enjoyment gap.

Linking these results back to Figure 3, Figure 4 depicts the fitted values predicted from Columns 2 (left), 4 (center), and 6 (right) in Table 5. Dots correspond to the conditional averages plotted in Figure 3. Different colors indicate the conditioning levels of realized enjoyment. The results show the quadratic and cubic specifications deliver almost identical predictions and that they fit the conditional averages much better than the linear model. More specifically, they show a steeper slope for negative values of the enjoyment gap but a rather flat shape for positive values, a first indication that losses in the enjoyment gap may be valued differently than gains.

Further confirming the different elasticities for positive (above expectations) and

FIGURE 4. LINEAR, QUADRATIC, AND CUBIC REGRESSIONS FOR THE ENJOYMENT GAP



Note: Solid lines depict the fitted values predicted from Columns 2 (left), 4 (center), and 6 (right) in Table 5. Circles correspond to the conditional averages shown in Table 5. Colors indicate the level of realized enjoyment. Shaded areas indicate 95% confidence intervals for the predictions.

TABLE 6—REGRESSIONS FOR ENJOYMENT BELOW, AT PAR, AND ABOVE EXPECTATIONS

	(1)	(2)	(3)
Above expectations	-0.000 (0.049)	0.022 (0.049)	0.013 (0.048)
Below expectations	-0.111 (0.044)	-0.124 (0.039)	-0.125 (0.039)
Controls	No	Yes	Yes
Night fixed effects	No	No	Yes
Adjusted R-squared	0.015	0.145	0.174
Num. of observations	433	433	433

Note: The table reports the regression coefficients of dummy variables indicating whether the reported realized enjoyment is above or below the reported expectation (the base category is at par with expectations). When included, controls coincide with those in Table 4. Standard errors, clustered by joint-payment groups, are reported in parentheses.

negative (below expectations) values of enjoyment gap, Table 6 shows the results of a regression of log payment on dummy variables for a positive or negative enjoyment gap, without controls (Column 1), with controls (Column 2), and with added night fixed effects (Column 3). Estimated coefficients indicate that compared with having an experience that met one’s expectations, having an experience that falls short of expectations negatively and significantly affects payment, whereas an experience that exceeds expectations has no significant effect on payments. Combined, our results support the prediction that being disappointed with respect to expectations has a greater impact on payments than being pleasantly surprised.

Finally, we run additional regressions of log payment on other satisfaction measures obtained in the ex-post questionnaire. Doing so is important because the extent to which expectations were met and the likelihood of recommending the show to others can be understood as single variables already capturing the mag-

TABLE 7—REGRESSION RESULTS FOR ALTERNATIVE MEASURES OF REPORTED ENJOYMENT

	Enjoyment measure	Adj. $R^2$	Num. obs.
Expectations were met	0.088 (0.037)	0.016	433
+ controls	0.116 (0.037)	0.156	433
+ night fixed effects	0.121 (0.033)	0.188	433
Likelihood of recommendation	0.102 (0.033)	0.031	434
+ controls	0.109 (0.029)	0.162	434
+ night fixed effects	0.112 (0.028)	0.194	434
Agreement with the message	0.043 (0.020)	0.013	419
+ controls	0.047 (0.021)	0.140	419
+ night fixed effects	0.047 (0.023)	0.172	419

*Note:* The table reports the regression coefficients of the variables indicated in the first row of the first column of each panel. Specifications labeled ”+ controls” include controls, and those indicated by ”+ night fixed effects” include controls and night fixed effects. When included, controls coincide with those in Table 4. Standard errors, clustered by joint-payment groups, are reported in parentheses.

nitude of the enjoyment gap; they can be used when obtaining both measures separately is not possible. Consistent with the results reported in Table 6, these two variables (Table 7) have positive and statistically significant coefficients in the corresponding regressions. The estimated values and the  $R^2$  are similar to those obtained for the enjoyment gap (Table 4). We find that the degree of agreement with the show’s message is a noisy indicator of payment, as expected.

Consistent with our predictions, the results of the field experiment confirm that payments increase with realized enjoyment, and that they further increase with the enjoyment gap. Importantly, we find that once satisfaction gap is accounted for, the effect of enjoyment on payment becomes non-significant. Finally, our findings show that, consistent with loss aversion, payments exhibit a steeper slope for negative enjoyment gaps than for positive enjoyment gaps of the same magnitude.

## V. Conclusion

Using PWYW pricing as a proxy for the utility of subjective consumption allowed us to unpack the factors influencing it. We focused our investigation on the interaction between expected and realized enjoyment as a key determinant of



the consumption utility. Our main finding is that the gap between participants' expectations and actual enjoyment is the main driver of payment. Controlling for this enjoyment gap, we find low expectations are not necessarily associated with greater enjoyment, and that people who report enjoying the show more do not necessarily pay more. The result remains robust after we control for several variables. In addition, payments exhibit loss-aversion patterns, with a steeper slope for negative values of the enjoyment gap than for positive values.

The present study offers insights useful for the marketing of experience goods. First, they highlight the important trade-off between setting expectations high enough to attract an audience and the risk that such expectations will result in disappointment, leading to lower payments or a decreased probability of customers returning. Further, disappointed consumers are likely to spread negative word of mouth, which can have detrimental effect on the product's success. Second, measuring consumer satisfaction with surveys answered post consumption does not capture the full picture since these surveys measure only the self-reported enjoyment but not the enjoyment gap. Third, our results highlight that pricing experience goods is particularly difficult and offer a potential alternative. Under traditional pricing mechanisms, fixed prices are paid before consumers actually know how much they will like the product. This situation can create two types of errors: buying a product that turns out to be a disappointment, or not buying one that would have exceeded expectations. PWYW pricing avoids these shortcomings, which may be a reason for why it is increasingly being used for several types of experience goods, such as software, music albums, restaurant meals, and pictures taken at touristic spots.

## REFERENCES

- Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman,** "Reference Points and Effort Provision," *American Economic Review*, April 2011, *101* (2), 470–492.
- Azar, Ofer H.,** "The Social Norm of Tipping: A Review," *Journal of Applied Social Psychology*, February 2007, *37* (2), 380–402.
- Balafoutas, Loukas, Adrian Beck, Rudolf Kerschbamer, and Matthias Sutter,** "What drives taxi drivers? A field experiment on fraud in a market for credence goods.," *Review of Economic Studies*, July 2013, *80* (3), 876–891.
- Bushong, Benjamin and Tristan Gagnon-Bartasch,** "Misattribution of Reference Dependence: Evidence from Real-Effort Experiments," mimeo, Harvard Business School, January 2017.

- Card, David E. and Gordon B. Dahl**, “Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior,” *Quarterly Journal of Economics*, February 2011, *126* (1), 103–143.
- Darby, Michael R. and Edi Karni**, “Free Competition and the Optimal Amount of Fraud,” *Journal of Law and Economics*, April 1973, *16* (1), 67–88.
- Dulleck, Uwe, Rudolf Kerschbamer, and Matthias Sutter**, “The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation, and Competition,” *American Economic Review*, April 2011, *101* (2), 526–555.
- Gneezy, Ayelet, Uri Gneezy, and Dominique O. Lauga**, “A Reference-Dependent Model of the Price–Quality Heuristic,” *Journal of Marketing Research*, April 2014, *51* (2), 153–164.
- , – , **Leif D. Nelson, and Amber Brown**, “Shared Social Responsibility: A Field Experiment in Pay-What-You-Want Pricing and Charitable Giving,” *Science*, July 2010, *329* (5989), 325–327.
- Kahneman, Daniel and Amos Tversky**, “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica*, March 1979, *47* (2), 263–292.
- and **Richard H. Thaler**, “Anomalies: Utility Maximization and Experienced Utility,” *Journal of Economic Perspectives*, Winter 2006, *20* (1), 221–234.
- , **Peter P. Wakker, and Rakesh Sarin**, “Back to Bentham? Explorations of Experienced Utility,” *Quarterly Journal of Economics*, May 1997, *112* (2), 375–406.
- Kőszegi, Botond and Matthew Rabin**, “A Model of Reference-Dependent Preferences,” *Quarterly Journal of Economics*, November 2006, *121* (4), 1133–1165.
- Lynn, Michael**, “Tipping in Restaurants and Around the Globe: An Interdisciplinary Review,” in Morris Altman, ed., *Handbook of Contemporary Behavioral Economics: Foundations and Developments*, New York: Taylor and Francis Group, 2006, chapter 31, pp. 626–645.
- Medvec, Victoria Husted, Scott F. Madey, and Thomas Gilovich**, “When less is more: Counterfactual thinking and satisfaction among Olympic medalists,” *Journal of Personality and Social Psychology*, October 1995, *69* (4), 603–610.
- Nelson, Phillip**, “Information and Consumer Behavior,” *Journal of Political Economy*, March–April 1970, *78* (2), 311–329.

**Schwartz, Barry**, *The Paradox of Choice: Why More Is Less. How the Culture of Abundance Robs US of Satisfaction*, New York: Harper Collins Publishers, 01 2005.

**Thaler, Richard**, "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior & Organization*, March 1980, 1 (1), 39–60.

APPENDIX A: ADDITIONAL TABLES AND FIGURES

FIGURE A1. QUESTIONNAIRES (Original and English Translation)

A. Pre-Show Questionnaire (Catalan)

Voluntari número: <u>16</u>	Ubicació del voluntari: <u>16</u>
Data del estudi: <u>14-03-2015</u>	

Gènere:  Dona  Home Edat: 23

Professió/Ocupació del voluntari: Psicòloga

Número de vegades anteriors que ha vingut a un espectacle a la Sala Beckett: 2

Número de vegades que ha anat a un espectacle de teatre en l'últim any: 4

¿Com ha conegut l'espectacle? Radio TV  Xarxes Amics Companys de feina Familiars

¿Ha reservat vostè mateix les entrades?  Sí  NO

¿De quantes persones ve acompanyat? 0 1  2 3 4 5 6 7 8 9 10 +

En una escala de 1 (mínim) al 7 (màxim) ¿Quant espera gaudir del espectacle?

1 2 3 4  5 6 7

Indiqui el color de la pastilla A347-B que li ha tocat:

BLANCA TARONJA  NO M'HAN DONAT CAP

B. Pre-Show Questionnaire (English Translation)

- Volunteer number \_\_\_\_\_ Volunteer Location \_\_\_\_\_
- Study Date \_\_\_\_\_
- Gender \_\_\_\_\_ Age \_\_\_\_\_
- Occupation \_\_\_\_\_
- Number of times visited this theater in the past \_\_\_\_\_
- Number of times going to the theater in the past year \_\_\_\_\_
- How did you hear about this play?  Radio  TV  Online  Friends  Colleagues  Family members
- Did you make the ticket reservations? Yes \_\_\_ No
- Number of people joining you today 0 1 2 3 4 5 6 7 8 9 10 +
- On a scale from 1 (minimum) to 7 (maximum), how much do you expect to enjoy the show?
- Tell us the color of the pill you took: White \_\_\_ Orange \_\_\_ I did not get any

C. Post-Show Questionnaire (Catalan)

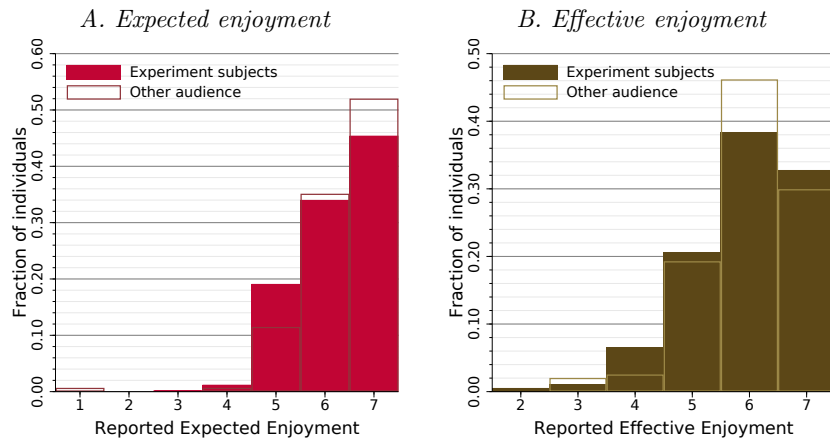
Voluntari número: <u>5</u>	Ubicació del voluntari: <u>5</u>
Data del estudi: <u>14-03-2015</u>	

- Creu que recomanarà aquest espectacle a unes altres persones?  
No    Pot ser    Probablement    **Segur que sí**
- En una escala d'1 (mínim) al 7 (màxim) Quant ha gaudit de l'espectacle?  
1    2    3    4    5    6    **7**
- Ha satisfet l'espectacle les seves expectatives?  
No    Més o Menys    **Sí**    les ha superat
- S'ha pres la pastilla que li hem donat abans de començar l'espectacle?  
**Sí**    No me l'he pres    No m'han donat cap pastilla
- En una escala d'1 (mínim) al 7 (màxim) Quant estàs d'acord amb el missatge de la obra?  
1    2    3    4    5    6    **7**
- Algun altre comentari?  
\_\_\_\_\_  
\_\_\_\_\_

D. Post-Show Questionnaire (English Translation)

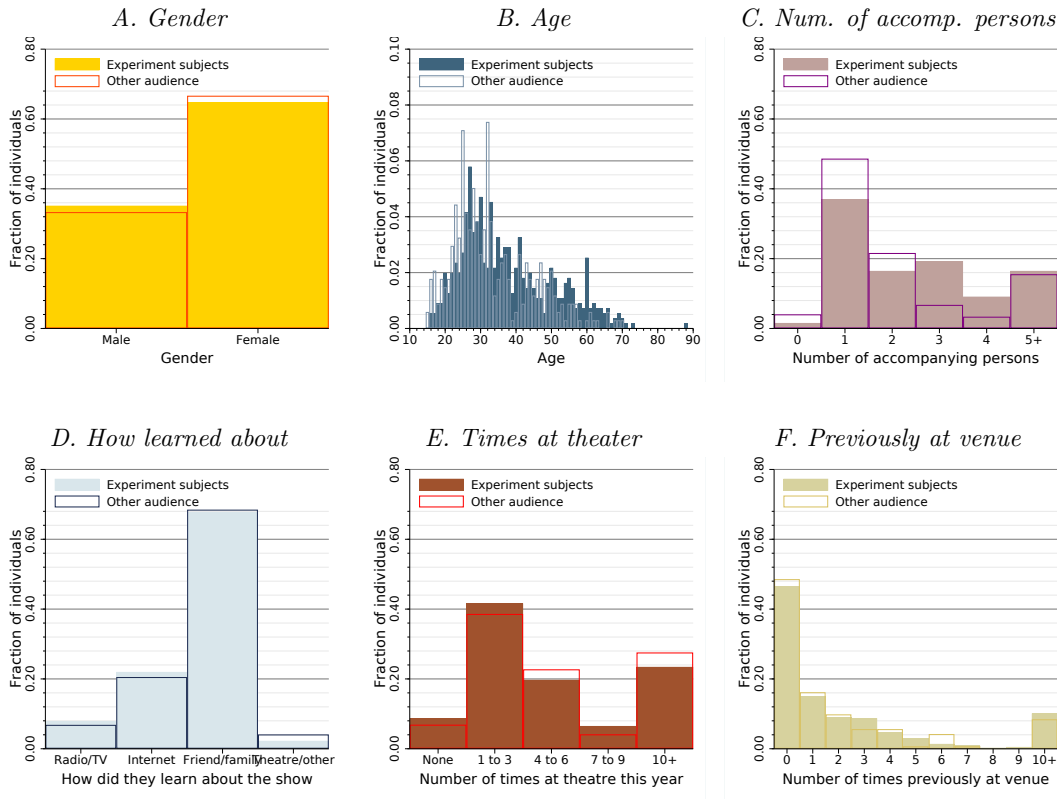
- Volunteer number \_\_\_\_\_ Volunteer Location \_\_\_\_\_
- Study Date \_\_\_\_\_
- Would you recommend this play to others?  No  Could be  Probably  For certain
- On a scale from 1 (minimum) to 7 (maximum), how much did you enjoy the show?
- Did the show meet your expectations?  No  More or less  Yes  It surpassed them
- Did you take the pill you were given before the show?  Yes  No  None given
- On a scale from 1 (minimum) to 7 (maximum), how much do you agree with the message of the play?
- Any other comments? \_\_\_\_\_

FIGURE A2. DISTRIBUTION OF REPORTED EXPECTED AND EX-POST ENJOYMENT (Participants Versus Other Audience)



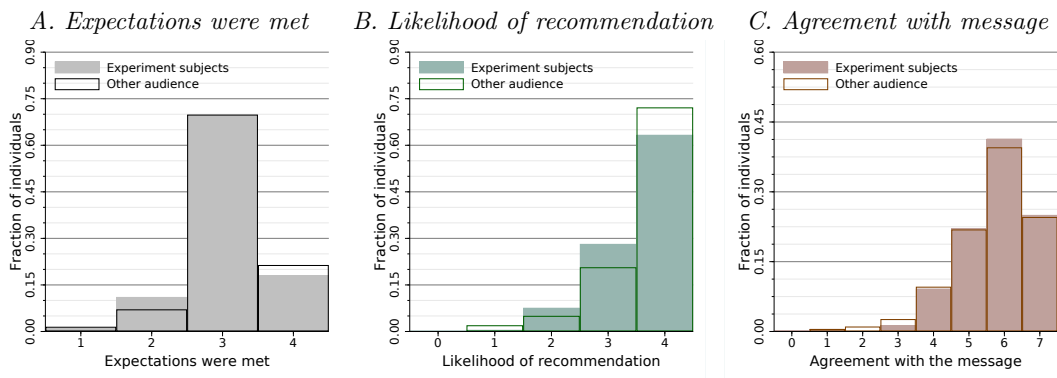
*Note:* The figure plots reported expected and realized enjoyment histograms for experiment participants (solid) and other audience (hollow), which comprises non-participants who completed the surveys during the last four nights of the show. Kolmogorov-Smirnov tests of the hypothesis that the two empirical distributions come from the same population distribution cannot reject the null hypothesis (p-values are 0.446 and 0.846, respectively).

FIGURE A3. DISTRIBUTIONS OF PERSONAL CHARACTERISTICS (Study Subjects Versus Other Audience)



*Note:* The figure plots the histograms of observable characteristics (gender, age, number of individuals in the group, how they heard about the show, number of times attended the theater this year, number of times visited the specific venue in the past) for participants (solid) and other audience (hollow). Kolmogorov-Smirnov tests of the hypothesis that the two empirical distributions for each characteristic come from the same population distribution cannot reject the null hypothesis except in the case of age and number of accompanying persons (p-values are 1.000, 0.000, 0.000, 1.000, 0.946, and 0.986, respectively).

FIGURE A4. DISTRIBUTION OF ALTERNATIVE MEASURES OF REPORTED ENJOYMENT (Study Subjects Versus Other Audience)



*Note:* The figure plots histograms of additional enjoyment-related measures (whether expectations were met, likelihood of recommendation, and agreement with the show’s moral) for participants (solid) and other audience (hollow). Kolmogorov-Smirnov tests of the hypothesis that the two empirical distributions come from the same population distribution cannot reject the null hypothesis (p-values are 0.992, 0.178, and 1.000, respectively).

TABLE A1—AVERAGE PAYMENT AND ENJOYMENT MEASURES FOR SELECTED GROUPS

	Payment	Expected enjoymt.	Realized enjoymt.	Expectat. met	Likelihd. of recom.	Agreemt. message
Age group:						
15-24	10.05 (0.52)	6.33 (0.10)	5.90 (0.15)	3.10 (0.09)	3.45 (0.10)	5.59 (0.17)
25-34	11.86 (0.31)	6.07 (0.07)	5.94 (0.08)	3.10 (0.05)	3.57 (0.05)	5.78 (0.09)
35-44	13.00 (0.49)	6.32 (0.07)	5.88 (0.09)	2.93 (0.05)	3.60 (0.06)	5.63 (0.11)
45-54	15.55 (0.76)	6.32 (0.12)	6.07 (0.11)	3.04 (0.06)	3.60 (0.08)	6.04 (0.11)
55+	15.17 (0.68)	6.15 (0.10)	5.85 (0.13)	2.98 (0.08)	3.45 (0.10)	5.63 (0.15)
Gender:						
Male	12.86 (0.38)	6.00 (0.07)	5.72 (0.09)	2.99 (0.05)	3.45 (0.06)	5.61 (0.09)
Female	12.81 (0.30)	6.32 (0.05)	6.04 (0.05)	3.07 (0.04)	3.60 (0.04)	5.80 (0.07)
Paid in group:						
Yes	13.32 (0.45)	6.15 (0.07)	5.88 (0.08)	3.05 (0.05)	3.54 (0.06)	5.78 (0.09)
No	12.52 (0.26)	6.24 (0.05)	5.96 (0.06)	3.03 (0.03)	3.55 (0.04)	5.70 (0.07)
Num. of accomp. persons:						
0	14.36 (2.75)	6.29 (0.29)	6.14 (0.40)	3.57 (0.20)	3.57 (0.30)	5.71 (0.42)
1-3	12.92 (0.34)	6.27 (0.06)	5.96 (0.07)	2.99 (0.04)	3.58 (0.05)	5.63 (0.09)
4-6	12.93 (0.66)	6.07 (0.10)	5.89 (0.12)	3.06 (0.08)	3.57 (0.07)	5.72 (0.14)
7-9	12.63 (0.59)	6.20 (0.08)	5.88 (0.10)	3.07 (0.05)	3.58 (0.07)	5.70 (0.11)
10+	11.80 (0.48)	6.23 (0.13)	6.10 (0.15)	3.03 (0.09)	3.46 (0.14)	5.85 (0.16)
First time at venue:						
Yes	13.27 (0.33)	6.28 (0.05)	5.84 (0.07)	2.97 (0.04)	3.46 (0.05)	5.64 (0.07)
No	12.32 (0.34)	6.11 (0.06)	6.03 (0.07)	3.11 (0.04)	3.65 (0.04)	5.83 (0.08)
First time this year:						
Yes	12.90 (0.24)	6.21 (0.04)	5.90 (0.05)	3.02 (0.03)	3.55 (0.03)	5.72 (0.06)
No	12.09 (0.97)	6.14 (0.14)	6.22 (0.15)	3.24 (0.11)	3.51 (0.11)	5.80 (0.17)
Satisfaction pill:						
Placebo	12.99 (0.34)	6.18 (0.06)	5.89 (0.07)	3.05 (0.04)	3.56 (0.05)	5.64 (0.08)
Treat- ment	12.61 (0.35)	6.20 (0.06)	5.94 (0.07)	3.02 (0.04)	3.54 (0.05)	5.80 (0.08)
No pill	13.19 (0.92)	6.44 (0.14)	6.12 (0.16)	3.08 (0.09)	3.54 (0.10)	6.00 (0.18)

*Note:* The table reports means and standard errors for the variables indicated in the top row of each column for the individuals with characteristics indicated in the first column.