AUA Working Paper Series No. 2014-4
July 2014

Detecting false positives in experimental auctions: A case study of projection bias in food consumption

Teresa Briz *[Corresponding author]*
Universidad Polit‼ecnica de Madrid
teresa.briz@upm.es

Andreas C. Drichoutis
Agricultural University of Athens
adrihout@aua.gr

Rodolfo M. Nayga, Jr.
University of Arkansas & Norwegian Agricultural Economics Research, Institute & Korea University
rnayga@uark.edu.

Agricultural **U**niversity of **A**thens ·
Department of Agricultural Economics
& Rural Development · http://www.aoa.aua.gr

Detecting false positives in experimental auctions: A case study of projection bias in food consumption

Teresa Briz*[1], Andreas C. Drichoutis†[2], and Rodolfo M. Nayga, Jr.‡[3]

[1]Universidad Politécnica de Madrid
[2]Agricultural University of Athens
[3]University of Arkansas & Norwegian Agricultural Economics Research Institute & Korea University

First Draft: July 2, 2014

Abstract: In this paper we argue that valuable information can be conveyed by looking at data coming from the training rounds of experimental auctions. As a case study, we use data from an experiment that seeks to elaborate on the mediating role of mood states on projection bias. Following a mood induction procedure, subjects are found to bid more under negative mood (as compared to positive mood) for products that are delivered in the future but bid less under negative mood for products that are delivered in present time. We show that if one had neglected insights gained from the training auction data, the researcher would have fallen prey to a case of a false positive result.

Keywords: projection bias, experimental auctions, type I error, false positive.

JEL Classification Numbers: C12, C90.

1 Introduction

Scientific hypothesis testing relies on methods of statistical inference to empirically establish that an effect is not due to chance alone. This has been the gold standard of science

*Departamento de Economía, E.T.S.Ingenieros Agrónomos, Universidad Politécnica de Madrid, Madrid, Spain, e-mail: teresa.briz@upm.es.

†Lecturer, Department of Agricultural Economics, Agricultural University of Athens, 11855, Athens, Greece, e-mail: adrihout@aua.gr.

‡Professor and Tyson Endowed Chair, Department of Agricultural Economics & Agribusiness, Division of Agriculture, University of Arkansas, Fayetteville, AR 72701, USA, and Adjunct Professor, Norwegian Agricultural Economics Research Institute and Korea University, e-mail: rnayga@uark.edu.

ever since Ronald A. Fisher's era. A 'test of significance' (Fisher, 1925) of a treatment effect establishes that the effect is statistically significant when the test statistic allows us to reject the null hypothesis of no difference between two conditions based on a pre-specified low probability threshold.

All statistical hypothesis tests have a probability of making one of two errors: an incorrect rejection of a true null hypothesis (type I error) representing a false positive; or a failure to reject a false null hypothesis (type II error) representing a false negative.[1] False positives have received a great deal of attention; academic journals are less likely to publish null results and p-value hacking makes false positives vastly more likely (Simmons et al., 2011). In addition, the pressure of using the criterion of statistical significance may have led published research to systematically overestimate effect sizes (Lane and Dunlap, 1978) and report inflated effects (Fanelli and Ioannidis, 2013).

Type II errors, on the other hand, have not been given similar attention. Zhang and Ortmann (2013) reviewed 95 papers published in *Experimental Economics* between 2010 and 2012 and found that only one article mentions statistical power and sample size issues. Replication studies (e.g., Maniadis et al., 2014) are particularly prone to false negatives because they are typically underpowered (Simonsohn, 2013).[2]

False positives may have more serious implications than false negatives by leading the research community into false avenues and wasting resources. The problem of false positives is further exacerbated by the fact that researchers not only file-drawer entire studies but also file-drawer subsets of analyses that produce non-significant results (Simonsohn et al., 2014). In addition, researchers rarely take the extra step of replicating their original study (for an exception see Kessler and Meier, 2014). Recently, Simonsohn et al. (2014) introduced p-curve (the distribution of statistically significant p-values for a set of independent findings) as a way to distinguish between selective reporting and truth. This approach overcomes limitations of previous approaches such as the "funnel plots" method (Duval and Tweedie, 2000; Egger et al., 1997), the "fail safe" method (Rosenthal, 1979; Orwin, 1983) and the "excessive-significance test" (Ioannidis and Trikalinos, 2007). The p-curve tool requires a set of studies to be included in the analysis and as such, single-papers should contain multiple studies and at least one direct replication of one of the studies (Simonsohn et al., 2014). Given that self-replication is rare in the literature, it is hardly possible to detect a false

[1] A type III error, typically not one that researchers often deal with, occurs when a researcher produces the right answer to the wrong question (Kimball, 1957). Kennedy (2002) warns that this is not to be confused with psychologists' type III error (Kaiser, 1960), which is concerned with concluding significance in the wrong direction.

[2] For example, Simonsohn et al. (2013) argue that the null result obtained in the replication study of Maniadis et al. (2014), is just a noisy estimate and that the *relative* effect size is comparable to the original study of Ariely et al. (2003).

positive from single-paper studies.

In this study we show that, in the context of experimental auctions, useful information can be conveyed by looking at the data coming from the training or practice rounds of single-paper studies. Experimental auctions have become a popular tool for applied economists to elicit people's willingness to pay (WTP) values due to their demand revealing properties. These auctions are considered demand revealing because of the (theoretically) incentive compatible nature of the auction mechanisms. However, experimental auctions are often unfamiliar to subjects. Consequently, most practitioners agree that employing a training phase prior to the actual valuation task is essential for subjects to abandon market-like heuristics such as "buying low" or for demonstrating the incentive compatibility of the auction.

While preceding an actual auction with practice rounds is common, bids from practice rounds are rarely recorded. Corrigan et al. (2014) note that a well-known economist said this about his work on experimental auctions, "I pulled up three old data sets associated with various published papers. Alas, it seems I did not enter the practice round data (normally with candy bars) for any of them." Indeed, it is uncommon in this literature to pay particular attention to practice/training rounds. Only a handful of papers have done so. For example, although Drichoutis et al. (2011) did not specifically analyze the bid data from the training auctions, they found that subjects with extensive training gave significantly higher bids in the real auctions that followed the practice rounds than minimally trained subjects. In another study, Corrigan et al. (2014) examined the relationship between practice and real bids from two auction experiments where participants bid on homegrown-value goods. They found a positive correlation between practice and real bids but that this was mitigated by repetition.

In this paper we use data from our study on projection bias in the context of experimental auctions.[3] We opted to look at the mediating effect of induced mood states on subjects' WTP for some products at three different delivery dates: i) in present time (right after the auction), ii) one week later and before typical lunch time, and iii) one week later and after typical lunch time. We used two different types of products, a ham-cheese sandwich for which craving at lunch time is relevant and a ballpoint pen for which lunch hours should not be relevant. Our results show that mood states actually mediate the effect of projection bias on subjects' WTP. However, a more careful look at the data coming from the training rounds indicates the presence of a similar effect for a set of (different) products used in the training rounds. Given that mood was induced only after the training rounds, we conclude that the effect we observe is not due to treatment assignment. If data from the training rounds have

[3] Loewenstein et al. (2003) coined the term "projection bias" to describe the general bias that has been documented in relation to the prediction of future tastes.

not been analyzed, our study would have contributed to the realm of false positive studies.

2 Methods

The experiment was carried out in February and March, 2013 at the Universidad Politécnica de Madrid. Announcements on the website of the University had been made a few weeks before the scheduled sessions. Students also received bulk announcement emails, sent to their university email accounts. The announcement did not provide specific details about the experiment, just general information that the experiment was about a study on consumer behavior. Participants responded with their weekday preferences and were then randomly assigned to one of the sessions. For each session, exactly eight participants were assigned with appropriate over-recruitment to account for no-shows.

The experiment consisted of 6 treatments in 24 different sessions (4 sessions/treatment). In all, 192 students participated in the experiment (8 subjects/session) and each subject only took part in one session. The experiment involved a 2 (mood inducement)$\times$3 (time of delivery of the product) between-subjects experimental design. With respect to mood inducement (described momentarily), half of the subjects were induced to a positive mood state and the other half were induced to a negative mood state. We also varied the time of delivery of the auctioned products to test the effect of projection bias. In one third of the sessions (8 sessions) the product was given right after the auction (control sessions); in another third of the sessions (8 sessions), the product was given one week later at 1 pm (which is typically considered "before lunch" in the Spanish culture and according to students' habits); and in the last third of the sessions (8 sessions), the product was given one week later at 3 pm (which is typically considered an "after lunch" time). We auctioned together one food and one non-food product to test the effect of craving on subjects' WTP. We would expect a priori that delivering the product before or after lunch time is relevant for the food item but is not relevant for the non-food item. For the "before lunch" and "after lunch" future delivery treatments, the highest bidders were given an exchangeable coupon and were told that a fresh sandwich bought on the delivery date would be available in the exact same place of the auction site one week later. The place where the auctions took place is one of the main classroom buildings in the center of campus, just a few meters from the head office building. In the control sessions, subjects were given the product right after the auction, which is the standard procedure for auctions. The experimental design is depicted in Table 1.

To achieve some variation over hunger levels, we varied the time of the sessions. Half of the sessions were conducted at 12 pm and half of the sessions were conducted at 1 pm. We

Table 1: Experimental design

		Mood	
		Negative	Positive
Time of delivery	Future at 12 pm	4 sessions × 8 subjects	4 sessions × 8 subjects
	Future at 3 pm	4 sessions × 8 subjects	4 sessions × 8 subjects
	Present	4 sessions × 8 subjects	4 sessions × 8 subjects

only allowed a one hour difference between scheduled sessions to minimize potential "time of the day" effects on bidding behavior (Demont et al., 2012, 2013; Hoffman et al., 1993; Menkhaus et al., 1992; Morawetz et al., 2011). Since we could not fully control subjects' eating behavior before participating in the auction, we also asked subjects to self-report their hunger level.

When subjects arrived at the lab they were randomly seated and were assigned a six-digit participant number. They were told that all their answers were confidential, that answers would only be used for this specific study and that they would be given €10 at the end of the session for their participation. We elicited subjects' WTP using the popular second-price auction (Vickrey, 1961). The experimenter carefully explained the auction mechanism by means of numerical examples that were also projected in a screen. Subjects then participated in hypothetical practice auctions to familiarize themselves with the procedure. In this training phase, subjects bid separately for a USB pendrive and a mug in three repeated rounds. The importance to bid their true value for the goods during the auction was emphasized to the subjects. At the end of the third round, one of the rounds and one of the products were chosen randomly as binding. Although subjects knew this (training) procedure was hypothetical and thus the binding round was not binding at all, we used this language to mimic as closely as possible the auction procedure of the real rounds. No information was posted between rounds.

Next, subjects were induced in either a positive mood state or a negative mood state. To induce subjects into different moods, we exposed them to picture stimuli. The stimuli consisted of 40 color pictures representing either pleasant or unpleasant scenes. Half of the subjects were induced to a positive mood state (exposed to pleasant pictures) and the other half were induced to a negative mood state (exposed to unpleasant pictures). Pictures were selected from the International Affective Picture System (IAPS) (Lang et al., 2008).[4] Each one of the pictures was shown for 6 seconds with a 10 second gap in between, in order to let

[4]The library numbers for IAPS pictures used in this study for positive mood inducement are: 1340, 1440, 1441, 1463, 1630, 1659, 1999, 2035, 2071, 2158, 2224, 2314, 2352, 2391, 2501, 2550, 2791, 4628, 5831, 8496; for negative mood inducement are: 1019, 2053, 2205, 2375, 2455, 2456, 2688, 2700, 2703, 3350, 6212, 6520, 8485, 9040, 9075, 9254, 9332, 9341, 9410, 9560.

participants rate their emotional experience on a 5-point Likert scale anchored by a 'smiley' face and a 'frowned' face (see experimental instructions in Appendix A).

To quantify the mood induction effect, we used the Positive and Negative Affect Scale (PANAS), developed by Watson et al. (1988). This scale consists of 20 items using 5-point scales (1 = very slightly/not at all to 5 = extremely). The scale is sub-divided in two 10-item scales for positive affect (PA) and negative affect (NA). The terms comprising each sub-scale for negative affect are: afraid, scared, nervous, jittery, irritable, hostile, guilty, ashamed, upset, distressed; while the terms for positive affect are: active, alert, attentive, determined, enthusiastic, excited, inspired, interested, proud and strong. We used the Spanish version of the PANAS scale validated by Robles and Páez (2003).

Following mood induction, we carried out the real (non-hypothetical) auction. The auctioned products were a ballpoint pen and a ham-and-cheese sandwich. Before the auction, subjects were able to examine visually the auctioned products: a standard black ballpoint pen, and a non-branded ham-and-cheese packed sandwich. We explained that the auction procedure was the same as the one used for the practice auction. Depending on treatment assignment, subjects were told that they would get the product i) right after the auction, ii) one week later at 12 pm, iii) one week later at 3 pm. In all treatments (including the future delivery treatments), the highest bidders had to pay the 2nd highest price for the product that they won right after the auction. Subjects were informed about this policy before they started bidding. Subjects bid in three repeated rounds with no information being posted in between rounds. At the end of the third round, one round and one product were chosen as binding.

At the end of the experiment, subjects signed a participation sheet and were given €10 (minus the 2nd highest price, if the person was the highest bidder in a binding round/product). In the present delivery treatment, subjects were also given the product they paid for, while in the future delivery treatments subjects were given a coupon to be redeemed a week later.

3 Results

3.1 Picture stimuli

We first explore whether subjects rated the emotional experience of the picture stimuli consistently with a priori expectations. Subjects in the positive mood treatment were exposed to a series of 20 pleasant pictures while subjects in the negative mood treatment were exposed to a series of 20 unpleasant pictures. They were then asked to indicate how each

picture made them feel on a 5-point Likert scale anchored by a 'smiley' face and a 'frowned' face. We then summed their responses over the 20 pictures.

Figure 1 depicts kernel density estimators of picture evaluation scores by treatment (present delivery, future delivery at 1 pm, future delivery at 3 pm). Vertical lines depict median values. It is evident that pleasant pictures were scored lower while unpleasant pictures were scored higher across all treatments. Note that the smiley scale was reversed, so that a lower score indicates feelings associated with a 'smiley' when watching the picture while a higher score indicates feelings associated with a 'frown'. A t-test of whether pleasant and unpleasant picture evaluation scores are significantly different in each treatment, highly reject the null. Similar results are obtained using non-parametric Wilcoxon-Mann-Whitney tests.

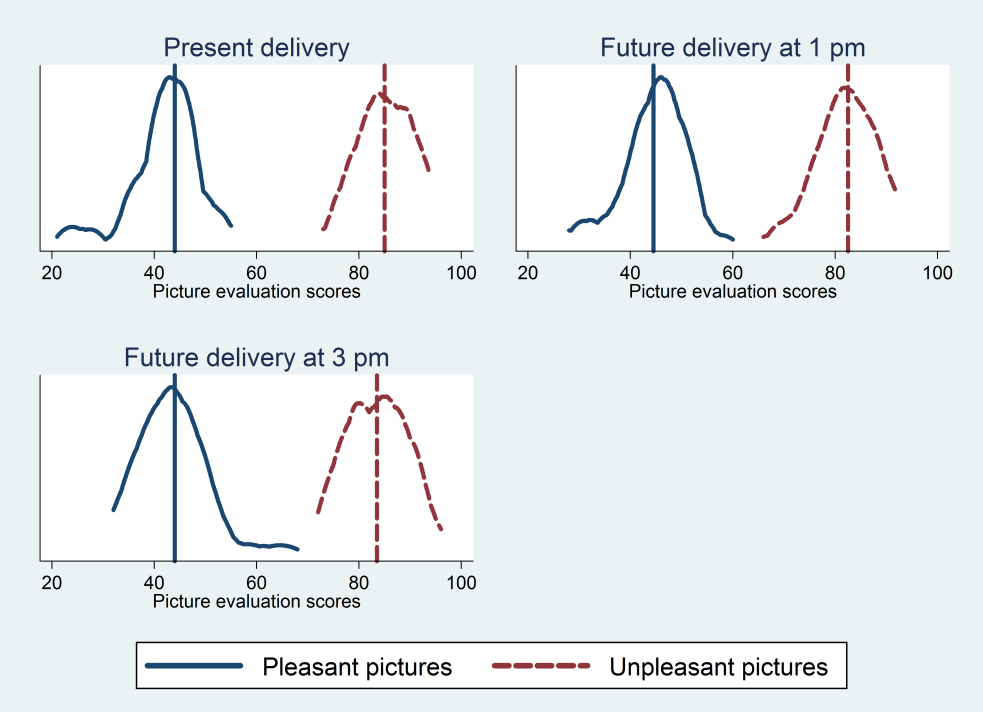


Figure 1: Kernel density estimators of picture evaluation scores by treatment

In addition, pleasant and unpleasant pictures were evaluated similarly across treatments. An ANOVA test indicates that evaluation scores for pleasant pictures do not differ across treatments (F-statistic=1.44, p-value=0.241). The same goes for unpleasant picture evaluation scores (F-statistic=2.41, p-value=0.096). We obtain similar results using the non-parametric Kruskal-Wallis test for pleasant pictures ($\chi^2 = 2.509$, p-value=0.285) and for unpleasant pictures ($\chi^2 = 3.997$, p-value=0.135), respectively. Figure B.1 in Appendix B depicts kernel density estimates of pleasant picture evaluation scores and unpleasant picture

evaluation scores. The graphs confirm results from the statistical tests above. In addition, Kolmogorov-Smirnov tests indicate that we cannot reject the null of equality of distributions of the picture evaluation scores. All in all, results from the tests above indicate that the set of pleasant pictures corresponded to a pleasant emotional experience while the set of unpleasant pictures corresponded to an unpleasant emotional experience.

3.2 Picture stimuli and induced mood

Once we established that subjects perceived pleasant (unpleasant) picture stimuli as pleasant (unpleasant), the next important question is whether the picture stimuli were adequate in inducing positive and negative mood states. For this purpose we summed the individual items of the PANAS scale to form the two sub-scales of positive and negative affect.

Figure 2 displays kernel density estimates for the positive and negative affect scales. The top panel corresponds to subjects in the positive mood treatment (these are the subjects that were exposed to pleasant picture stimuli) while the lower panel corresponds to subjects in the negative mood treatment (these are the subjects that were exposed to unpleasant picture stimuli). Vertical lines depict median values. In the positive mood treatment the distribution for the positive affect scale is shifted to the right (indicating higher positive affect) while the distribution for the negative affect scale is shifted to the left (indicating lower negative affect). Thus, it appears that the pleasant picture stimuli were adequate in generating a shift in affect scores in the expected direction. Let us now contrast the top panel of Figure 2 with the bottom panel in the same figure. Compared with the positive mood treatment, positive affect is shifted to the left (implying lower levels of positive affect), while negative affect is shifted to the right (implying higher levels of negative affect). Thus, the unpleasant picture stimuli were also adequate in shifting the distribution of positive and negative affect in the expected directions.

Visual differences in Figure 2 are also confirmed by statistical analysis. A t-test of whether positive affect scores is different than negative affect scores, highly rejects the null in the positive mood inducement (p-value$<$0.001) as well as in the negative mood inducement (p-value=0.021). Results from Wilcoxon-Mann-Whitney tests agree with previous conclusions. Kolmogorov-Smirnov tests on whether the distributions of positive affect and negative affect are equal, highly reject the null in all cases. Additionally, one may wonder if induced positive affect levels are similar in the positive mood treatment and in the negative mood treatment. The evidence is clear: the t-tests, Wilcoxon-Mann-Whitney tests and Kolmogorov-Smirnov tests all indicate that the positive affect in the positive mood treatment is statistically

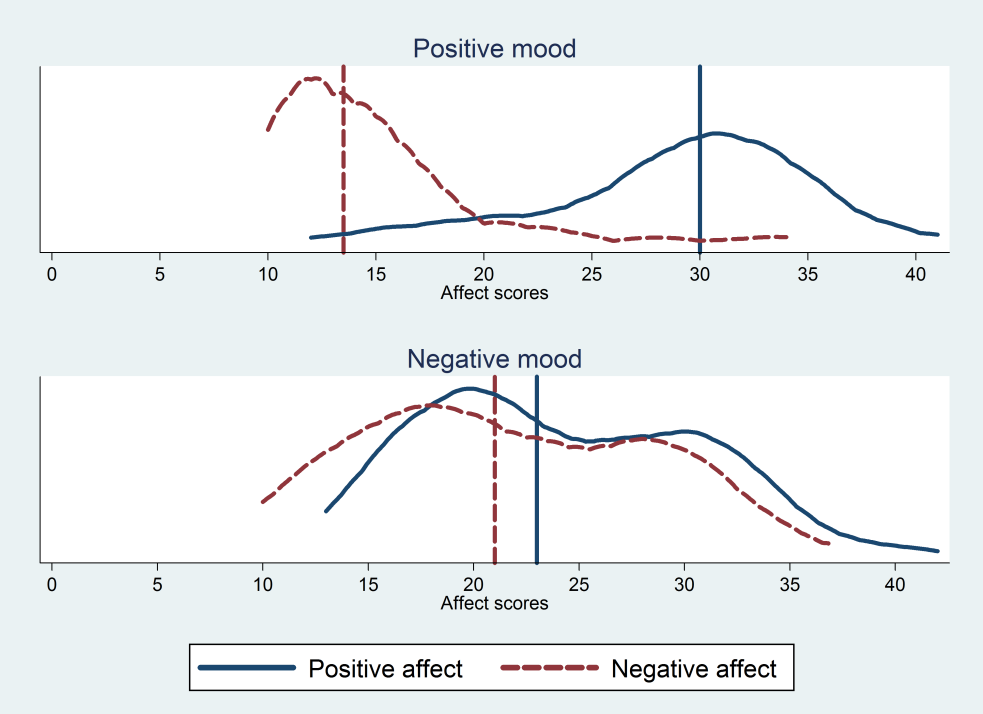


Figure 2: Kernel density estimators of positive and negative affect by mood induction treatment

different (and higher) than in the negative mood treatment. The same tests show that the negative affect in the positive mood treatment is statistically different (and lower) than in the negative mood treatment.

Overall, the results from this section boil down to this. The series of pleasant picture stimuli were adequate in increasing positive affect and decreasing negative affect in the positive mood treatment. The series of unpleasant picture stimuli were also adequate in decreasing positive affect and increasing negative affect in the negative mood treatment. It is important to note that while positive and negative affect levels are close in the negative mood treatment, this is not surprising given that positive and negative affect are aspects of mood that co-exist. All it matters is that the positive and the negative mood treatments exhibit affect levels in the expected directions and are different from each other.

3.3 Bidding behavior

Table 2 shows the mean, standard deviation and median values of bids (pooled over the three rounds) per product and treatment. The top two panels show the descriptive statistics for the USB and sandwich products. With respect to the sandwich product, we can see that subjects under negative mood bid more in both the future delivery treatments while they

bid less in the present delivery treatment (as compared to positive mood). There is a similar pattern for the USB with the exception that subjects do not bid differently in the future delivery at 3 pm treatment.

Table 2: Descriptive statistics of bids per treatment and product

Product	Treatment	Positive mood			Negative mood		
		Mean	SD	Median	Mean	SD	Median
Sandwich	Present delivery	1.52	0.77	1.50	1.27	0.66	1.20
	Future delivery at 1 pm	1.35	0.67	1.23	1.71	1.15	1.50
	Future delivery at 3 pm	0.98	0.84	0.83	1.17	0.85	1.00
Pen	Present delivery	1.05	0.73	1.00	0.82	0.47	0.75
	Future delivery at 1 pm	0.82	0.58	0.70	1.25	0.97	1.00
	Future delivery at 3 pm	0.75	0.76	0.50	0.74	0.81	0.50
USB pendrive	Present delivery	4.04	2.24	3.00	4.52	2.86	4.00
	Future delivery at 1 pm	5.07	3.16	5.00	7.74	5.11	7.00
	Future delivery at 3 pm	5.08	3.99	4.00	5.20	3.78	4.00
Mug	Present delivery	2.84	1.72	2.50	2.67	2.09	2.50
	Future delivery at 1 pm	2.78	1.73	2.25	4.94	3.89	3.75
	Future delivery at 3 pm	2.73	2.30	2.00	3.45	2.34	3.00

Notes: SD stands for standard deviation.

Although the analysis up to now is purely descriptive, a picture starts emerging. It looks as if mood states interact with time of delivery of the product which would imply that some form of interaction between projection bias and mood is at work. Table 3 shows the results from statistical tests of whether bids (per product and per time of delivery treatment) differ between positive and negative mood. Three sets of tests are presented: a t-test, a Wilcoxon-Mann-Whitney test (WMN), and a Kolmogorov-Smirnov (KS) equality of distributions test. The top two panels of Table 3 show that the difference in bids shown in Table 2 between the positive and negative mood treatments are statistically significant for the present and future delivery at 1 pm treatments. This result holds for both the sandwich and ballpoint pen under all three sets of tests. The statistical tests are not very clear for the sandwich product in the case of the future delivery at 3 pm treatment since the WMN test marginally rejects the null at the 5% level but fails to reject the null with a t-test and KS test.

At this point, we posit that most empirical researchers would be looking for narrative arguments to support the claim that there is some form of interaction between mood and projection bias as well as to explain how the small difference between the pen and sandwich results (in the future delivery at 3 pm treatment) could be attributed to craving. After all, the statistical tests look fairly convincing.

The bottom two panels in Table 2 show the descriptive statistics of bids from the practice

Table 3: Tests of differences in bids between the mood treatments

Product	Treatment	t-test		Wilcoxon-Mann-Whitney test		Kolmogorov-Smirnov test	
		t	p-value	z	p-value	D	p-value
Sandwich	Present delivery	-2.428	0.016	-2.025	0.043	0.167	0.106
	Future delivery at 1 pm	2.605	0.010	1.776	0.076	0.208	0.021
	Future delivery at 3 pm	1.549	0.123	1.959	0.050	0.156	0.151
Pen	Present delivery	-2.646	0.009	-1.536	0.125	0.240	0.005
	Future delivery at 1 pm	3.719	0.000	2.794	0.005	0.292	0.000
	Future delivery at 3 pm	-0.116	0.908	-0.327	0.743	0.063	0.988
USB pendrive	Present delivery	1.300	0.195	1.057	0.290	0.135	0.285
	Future delivery at 1 pm	4.357	0.000	4.004	0.000	0.365	0.000
	Future delivery at 3 pm	0.216	0.829	0.586	0.558	0.073	0.945
Mug	Present delivery	-0.628	0.531	-1.180	0.238	0.115	0.490
	Future delivery at 1 pm	4.975	0.000	4.090	0.000	0.260	0.002
	Future delivery at 3 pm	2.150	0.033	2.596	0.009	0.240	0.005

rounds for the USB pendrive and mug. Although these are data that researchers usually discard and (almost) never formally analyze, it is interesting to note that for both products subjects bid more in the negative mood treatments (as compared to the positive mood treatments) and future delivery treatments. Note that although we present the descriptive statistics per treatment for the two practice products, the treatments are irrelevant for the practice rounds. For one, in the practice rounds the language in the instructions did not mention any future delivery of the products nor were subjects made aware that there was a future delivery treatment in the later part of the experiment. Second, mood induction had not even been applied in the practice rounds nor were subjects made aware that they were going to be exposed to a series of images as part of a mood induction procedure. The bottom two panels in Table 3 present the statistical tests for the treatment effects shown in Table 2. These roughly support the discussion above.

Conceptually the problem with the analysis described above is that if a treatment effect is evident before the treatment is even applied, then the effect is likely not true. We argue that practice rounds in experimental auctions may convey useful information and data from practice rounds should always be analyzed before one goes and make big claims about treatment effects. At this point, one may argue that the reason we obtain significant results between the different treatments (even before we apply the treatment) is because the treatments differ (a lot) in terms of observables. This requires further investigation since we indeed reject the null of no difference for a few observable characteristics such as age (Kruskal-Wallis $\chi^2 = 26.72$, p-value< 0.001) and education level (Fisher's exact p-

value=0.001). However, we also fail to reject the null of no difference between treatments for gender ($\chi^2 = 5.19$, p-value=0.393), income level ($\chi^2 = 17.42$, p-value=0.294), hunger level (Kruskal-Wallis $\chi^2 = 8.35$, p-value=0.138), liking of ham-cheese sandwiches ($\chi^2 = 24.88$, p-value=0.206) and whether subjects had brought lunch with them the day of the experiment ($\chi^2 = 0.62$, p-value=0.987).

Naturally, we can control for observable characteristics in a regression context. Table 4 shows the results from random effects regressions separately for each product. Regression model for the sandwich product controls for extra factors that are relevant to the sandwich such as hunger level, liking of cheese-ham sandwich in general and whether the respondent brought lunch with him/her the day of the experiment. What is clear in Table 4 is that there are significant treatment effects consistent with our discussion above which, however, are also evident for the training products for which the treatments were not even applied. Note, that observable characteristics do not seem to play a significant role since regression results from models with just the treatment variable produce almost identical results (see Table 5).

4 Conclusion

Our results show the relevance of analyzing all available experimental data before making claims of significant effects. In the context of experimental auctions we argue that training rounds contain valuable information that researchers most often dismiss.

Using a case study on projection bias in experimental auctions, we demonstrate that data analysis from training rounds can shield against false positives. When we analyze the data from the real auction we detect significant effects that show that projection bias is mediated by mood. However, similar effects are present for the training rounds and products. Given that treatments were applied after the training auction, we conclude that the effect we observe is not due to assignment to treatment.

As a further step in this research agenda, it would be interesting to collect auction datasets from published papers for which data from the training rounds have been recorded. An analysis of data coming from the training rounds alongside the actual results reported in the published papers would be a first step in redirecting the research community from false avenues, if false positives have indeed been prevalent.

For future auction experiments we suggest that journal editors and reviewers ask authors to report data and analysis from the training phase of the experiment. If authors try to demonstrate effects from their experiments, these effects should not be present in the training part of the experiment, subject to the training part is performed before the treatment takes

Table 4: Random effects regressions (with demographics)

	(1) Sandwich		(2) Pen		(3) USB		(4) Mug	
Constant	-0.178	(0.595)	0.788	(0.518)	7.953***	(2.526)	2.711	(1.752)
Positive mood	0.327*	(0.190)	0.252	(0.180)	-1.282	(0.878)	-0.064	(0.609)
Future at 1 pm	0.620***	(0.187)	0.450***	(0.174)	3.191***	(0.848)	2.230***	(0.588)
Future at 3 pm	0.098	(0.185)	-0.036	(0.176)	0.249	(0.857)	0.674	(0.595)
Positive mood× Future at 1 pm	-0.873***	(0.270)	-0.686***	(0.253)	-1.463	(1.232)	-2.365***	(0.855)
Positive mood× Future at 3 pm	-0.565**	(0.265)	-0.219	(0.251)	1.189	(1.224)	-0.847	(0.849)
Round 2	-0.017	(0.028)	0.052*	(0.028)	-0.133	(0.134)	0.074	(0.078)
Round 3	0.021	(0.028)	0.064**	(0.028)	0.077	(0.134)	0.173**	(0.078)
Hunger_4	-0.249	(0.152)						
Hunger_3	0.051	(0.249)						
Hunger_2	0.369	(0.246)						
Hunger_1	-0.419	(0.263)						
Age	0.049**	(0.019)	0.005	(0.018)	-0.081	(0.088)	0.025	(0.061)
Undergrad student	0.113	(0.149)	0.129	(0.138)	-1.530**	(0.673)	-0.352	(0.467)
Male	-0.212*	(0.111)	-0.186*	(0.104)	-0.502	(0.505)	-0.054	(0.350)
Income_2	-0.296*	(0.179)	-0.200	(0.169)	-0.665	(0.822)	-0.403	(0.570)
Income_3	-0.104	(0.183)	-0.133	(0.173)	0.732	(0.842)	-0.456	(0.584)
Income_4	-0.236	(0.201)	-0.200	(0.185)	0.250	(0.904)	0.337	(0.627)
Did not bring lunch	-0.120	(0.115)						
Like sandwich_2	0.610***	(0.197)						
Like sandwich_3	0.840***	(0.189)						
Like sandwich_4	0.785***	(0.196)						
Like sandwich_5	0.885***	(0.213)						
σ_u	0.694***	(0.037)	0.672***	(0.036)	3.274***	(0.176)	2.289***	(0.121)
σ_ϵ	0.278***	(0.010)	0.269***	(0.010)	1.311***	(0.047)	0.764***	(0.028)
N	576		576		576		576	
Log-likelihood	-366.297		-347.810		-1259.616		-981.978	
AIC	782.594		727.620		2551.232		1995.955	
BIC	891.497		797.318		2620.930		2065.653	

Standard errors in parentheses. * p<0.1, ** p<0.05 *** p<0.01

Table 5: Random effects regressions

	(1) Sandwich		(2) Pen		(3) USB		(4) Mug	
Constant	1.264***	(0.143)	0.781***	(0.125)	4.542***	(0.616)	2.585***	(0.420)
Positive mood	0.251	(0.201)	0.234	(0.175)	-0.481	(0.865)	0.174	(0.590)
Future at 1 pm	0.441**	(0.201)	0.432**	(0.175)	3.216***	(0.865)	2.276***	(0.590)
Future at 3 pm	-0.097	(0.201)	-0.081	(0.175)	0.674	(0.865)	0.784	(0.590)
Positive mood× Future at 1 pm	-0.605**	(0.284)	-0.663***	(0.247)	-2.192*	(1.223)	-2.334***	(0.835)
Positive mood× Future at 3 pm	-0.439	(0.284)	-0.221	(0.247)	0.360	(1.223)	-0.893	(0.835)
Round 2	-0.017	(0.028)	0.052*	(0.028)	-0.133	(0.134)	0.074	(0.078)
Round 3	0.021	(0.028)	0.064**	(0.028)	0.077	(0.134)	0.173**	(0.078)
σ_u	0.786***	(0.042)	0.682***	(0.037)	3.374***	(0.181)	2.320***	(0.123)
σ_ϵ	0.278***	(0.010)	0.269***	(0.010)	1.311***	(0.047)	0.764***	(0.028)
N	576		576		576		576	
Log-likelihood	-389.226		-350.619		-1265.110		-984.437	
AIC	798.453		721.237		2550.221		1988.875	
BIC	842.014		764.798		2593.782		2032.436	

Standard errors in parentheses. * p<0.1, ** p<0.05 *** p<0.01

place.

References

Ariely, D., G. Loewenstein, and D. Prelec (2003). "coherent arbitrariness": Stable demand curves without stable preferences. *Quarterly Journal of Economics 118*(1), 73–105. 10.1162/00335530360535153.

Corrigan, J. R., M. C. Rousu, and D. P. T. Depositario (2014). Do practice rounds affect experimental auction results? *Economics Letters 123*(1), 42–44.

Demont, M., P. Rutsaert, M. Ndour, W. Verbeke, P. A. Seck, and E. Tollens (2013). Experimental auctions, collective induction and choice shift: willingness-to-pay for rice quality in Senegal. *European Review of Agricultural Economics 40*(2), 261–286.

Demont, M., E. Zossou, P. Rutsaert, M. Ndour, P. Van Mele, and W. Verbeke (2012). Consumer valuation of improved rice parboiling technologies in Benin. *Food Quality and Preference 23*(1), 63–70.

Drichoutis, A. C., J. Rodolfo M. Nayga, and P. Lazaridis (2011). The role of training in experimental auctions. *American Journal of Agricultural Economics 93*(2), 521–527.

Duval, S. and R. Tweedie (2000). Trim and fill: A simple funnel-plotbased method of testing and adjusting for publication bias in meta-analysis. *Biometrics 56*(2), 455–463.

Egger, M., G. D. Smith, M. Schneider, and C. Minder (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ 315*(7109), 629–634.

Fanelli, D. and J. P. A. Ioannidis (2013). US studies may overestimate effect sizes in softer research. *Proceedings of the National Academy of Sciences*.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

Hoffman, E., D. J. Menkhaus, D. Chakravarti, R. A. Field, and G. D. Whipple (1993). Using laboratory experimental auctions in marketing research: A case study of new packaging for fresh beef. *Marketing Science 12*(3), 318–338.

Ioannidis, J. P. A. and T. A. Trikalinos (2007). An exploratory test for an excess of significant findings. *Clinical Trials 4*(3), 245–253.

Kaiser, H. F. (1960). Directional statistical decisions. *Psychological Review 67*(3), 160–167.

Kennedy, P. E. (2002). Sinning in the basement: What are the rules? The ten commandments of applied econometrics. *Journal of Economic Surveys 16*(4), 569–589.

Kessler, J. B. and S. Meier (2014). Learning from (failed) replications: Cognitive load manipulations and charitable giving. *Journal of Economic Behavior & Organization 102*(0), 10–13.

Kimball, A. W. (1957). Errors of the third kind in statistical consulting. *Journal of the American Statistical Association 52*(278), 133–142.

Lane, D. M. and W. P. Dunlap (1978). Estimating effect size: Bias resulting from the significance criterion in editorial decisions. *British Journal of Mathematical and Statistical Psychology 31*(2), 107–112.

Lang, P. J., M. M. Bradley, and B. N. Cuthbert (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical report A-8, University of Florida, Gainesville, FL.

Loewenstein, G., T. O'Donoghue, and M. Rabin (2003). Projection bias in predicting future utility. *The Quarterly Journal of Economics 118*(4), 1209–1248.

Maniadis, Z., F. Tufano, and J. A. List (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *American Economic Review 104*(1), 277–90.

Menkhaus, D. J., G. W. Borden, G. D. Whipple, E. Hoffman, and R. A. Field (1992). An empirical application of laboratory experimental auctions in marketing research. *Journal of Agricultural and Resource Economics 17*(1), 44–55.

Morawetz, U. B., H. De Groote, and S. C. Kimenju (2011). Improving the use of experimental auctions in africa: Theory and evidence. *Journal of Agricultural and Resource Economics 36*(2), 263–279.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics 8*(2), 157–159.

Robles, R. and F. Páez (2003). Estudio sobre la traducción al español y las propiedades psicométricas de las escalas de afecto positivo y negativo (PANAS). *Salud Mental 26*(1), 69–75.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin 86*(3), 638–641.

Simmons, J. P., L. D. Nelson, and U. Simonsohn (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science 22*(11), 1359–1366.

Simonsohn, U. (2013). Small telescopes: Detectability and the evaluation of replication results. *Working paper, Available at http://dx.doi.org/10.2139/ssrn.2259879*.

Simonsohn, U., L. D. Nelson, and J. P. Simmons (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General 143*(2), 534–547.

Simonsohn, U., J. P. Simmons, and L. D. Nelson (2013). Anchoring is not a false-positive: Maniadis, tufano, and list's (2014) 'failure-to-replicate' is actually entirely consistent with the original. *Working paper, Available at http://dx.doi.org/10.2139/ssrn.2351926*.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance 16*(1), 8–37.

Watson, D., L. A. Clark, and A. Tellegen (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology 54*(6), 1063–1070.

Zhang, L. and A. Ortmann (2013). Exploring the meaning of significance in experimental economics. *Australian School of Business Research Paper No. 2013 ECON 32*.

A Appendix: Experimental Instructions

[This is an English translation of the original instructions written in Spanish]
[Text in brackets was not shown to subjects]

Welcome announcement
Thank you for agreeing to participate in this survey. The survey concerns the economics of decision making.

In a short while, we will conduct a series of experimental auctions with known products.

You have been randomly assigned a participant number for this entire session. All information collected is strictly confidential and will only be used for this specific project.

To thank you for your participation, you are going to be given 10€ at the end of the session. At the auction, you will have the opportunity to bid on and get the product, which will be given to the highest bidders (according to the rules I will describe momentarily).

If you have any questions, you may ask an assistant or the moderator. Do not communicate with other participants of this session.

In this experiment we explore several topics; therefore you will participate in several phases.

Page **1** out of **8**

------Page break------

The 3rd Price Vickrey Auction

In the tasks to follow you will participate in a type of auction known as a 3rd price auction. The 3rd price auction has 5 basic steps:

Step 1: We'll describe to you the product to be auctioned.

Step 2: Each one of you, will submit a bid for buying the product.

Step 3: The monitor will collect the bid sheets and rank all bids from highest to lowest.

Step 4: The persons that submit bids higher than the 3rd highest price buy the product **but will pay the price of the third highest bidder**. If your bid is not higher than the third highest bid then you dont purchase the good.

Consider this numerical example:

Suppose 8 people bid in an auction in order to buy a USB memory stick (16GB). Each bidder submits a bid separately. The submitted bids are given in the table below:

Person	**Bid**
1	12
2	15
3	20
4	18
5	30
6	25
7	35
8	0

------Page break------

After ranking bids from highest to lowest, we have:

Person	**Bid**
7	35
5	30
6	**25**
3	20
4	18
2	15
1	12
8	0

Persons 7 and 5 purchase one unit of the good because s/he bid higher than the 3rd highest price (35 and 30 respectively) but only pay 25 (third highest bid). All the other participants in the auction pay nothing and do not receive a memory stick.

In this auction, the best strategy is to bid exactly what the item is worth to you. Consider the following: if you bid less than what the object is worth to you, then you may not buy the product and miss a good opportunity for buying something at a price you were actually willing to pay.

Conversely, if you bid more than what the object is worth to you, you may end up having to pay a price higher than what you really wanted to. Thus, your best strategy is to bid exactly what the object is worth to you. The tasks you will do today are not hypothetical

and have real monetary consequences.

Do you have any questions?

------Page break------

Training auction [Hypothetical: USB memory stick]

We will now do a training task. This task is designed to allow you to familiarize yourself with the 3rd price auction. We will repeat this auction for three rounds. We will then select one round as binding by having one of you selecting a number from 1 to 3 from an urn. The numbers correspond to rounds, so if s/he picks number 1 then round 1 is binding, if s/he picks number 2 then round 2 is binding etc.

In this auction we will auction a memory stick. Take a look at this picture.

[Experimenter shows picture of usb memory stick in the screen]

You have all been provided with yellow slips, wherein you will write down and record your bid.

Please, take the yellow slip number 1 and write down the maximum you are willing to pay to purchase this usb stick.

After you've finished writing your bids, the monitor will go around the room and collect the bid sheets. I will then rank bids from highest to lowest, determine the 3rd highest price and the persons with bids above the 3rd highest price. In private, at the front of the room, bids will be ranked from lowest to highest.

The bid is private information and should not be shared with anybody else. Please be quiet while the auction is carried out.

[Once the first round is finished, second round starts]

Now, please, take the yellow slip number 2 and write down the maximum you are willing to pay to purchase this usb stick.

[Same procedure is followed. Once the second round is finished, third round starts]

Now, please, take the yellow slip number 3 and write down the maximum you are willing to pay to purchase this usb stick.

[Experimenter collects bid sheets]

[The experimenter asks one person to draw a number from an urn; Number determines binding round]

[IDs of highest bidders and 3rd highest price for the binding round are determined and announced]

Page **4** out of **8**

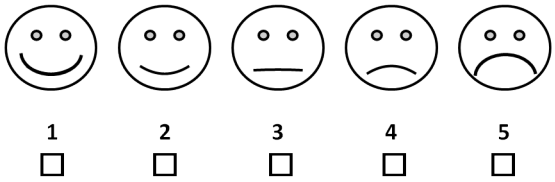
------Page break------

Picture evaluation phase [Mood inducement phase]

In this phase we will show you a sequence of 20 pictures.

You will see a slide with the number of the picture for 2 seconds, after it, each picture will be shown on the screen for 6 seconds, and then you have 10 seconds to describe "how you felt while watching the picture." Please, look at the pictures carefully and keep quiet.

To describe how you felt, we have provided a scale with faces that you should use for ranking.



Please, make sure the number of the picture matches the number of scale used and tick the appropriate box.

We remind you to please remain silent during the whole session.

Page **5** out of **8**

------Page break------

Feelings evaluation [Mood measurement]

This question consists of a number of words and phrases that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word.

Indicate to what extent you feel like this right now. Use the following scale to record your answers:
[Original word in Spanish is provided in parenthesis]

very slightly/not at all	a little	moderately	quite a bit	extremely
1	2	3	4	5

___	Interested (Motivado)	___	Irritable (Irritable)
___	Upset (Molesto a disgusto)	___	Alert (Alerta)
___	Excited (Emocionado)	___	Ashamed (Avergonzado)
___	Distressed (De malas)	___	Inspired (Inspirado)
___	Strong (Firme)	___	Nervous (Nervioso)
___	Guilty (Culpable)	___	Determined (Decidido)
___	Scared (Temeroso)	___	Attentive (Estar atento)
___	Hostile (Agresivo)	___	Jittery (Inquieto)
___	Enthusiastic (Entusiasmado)	___	Active (Activo)
___	Proud (Estar orgulloso)	___	Afraid (Inseguro)

------Page break------

The real product auctions

[Present auction treatment]

We will now auction a pen and a non-branded ham and cheese sandwich. The sandwiches are kept in a refrigerator and have been bought this morning. We will follow the same procedure as we did for the memory stick. We will repeat this auction for three rounds. We will then select one round as binding by having one of you selecting a number from 1 to 3 from an urn. The numbers correspond to rounds, so if s/he picks number 1 then round 1 is binding, if s/he picks number 2 then round 2 is binding etc. Finally, we will select one product as binding. We will select either 1 or 2 from an urn, being 1 the pen and 2 the ham and cheese sandwich. Please, pass the sandwiches and the pens around.

Please, take the pink slip number 1 and write down the maximum you are willing to pay to purchase the sandwich and the pen, respectively.

After you've finished writing your bids, the monitor will go around the room and collect the bid sheets. The monitor will then rank bids from highest to lowest, determine the 3rd highest price and the persons with bids above the 3rd highest price for each product, respectively.

The bid is private information and should not be shared with anybody else. Please be quiet while the auction is carried out.

[Once the first round is finished, second round starts]

Now, please, take the pink slip number 2 and write down the maximum you are willing to pay to purchase the sandwich and the pen, respectively.

[Same procedure is followed. Once the second round is finished, third round starts]

Now, please, take the pink slip number 3 and write down the maximum you are willing to pay to purchase the sandwich and the pen, respectively.

[Experimenter collects bid sheets]

[The experimenter asks one person to draw a number from an urn; Number determines binding round. The experimenter asks one person to draw a number from an urn. The number determines the binding product.]

[IDs of highest bidders and 3rd highest price for the binding product and round are determined and announced.]

[**Future auction treatment**]

We will now auction a pen and a non-branded ham and cheese sandwich. We will follow the same procedure as we did for the memory stick. We will repeat this auction for three rounds. We will then select one round as binding by having one of you selecting a number from 1 to 3 from an urn. The numbers correspond to rounds, so if s/he picks number 1 then round 1 is binding, if s/he picks number 2 then round 2 is binding etc. Finally, we will select one product as binding. We will select either 1 or 2 from an urn, being 1 the pen and 2 the ham and cheese sandwich. Please, pass the sandwiches and the pens around.

[**Before lunch treatment**]

The selected product will be given at 1 pm, in this exact place in one week from today. In case the binding product is a sandwich, we will have available fresh sandwiches made on the day of delivery. To ensure you get your pen or sandwich next week, you will be given a coupon to be redeemed, and my own professional card, in case you have any problem showing up. You can stop by my office to get the product in a weeks' time, if for some reason you are not able to pick it up from this room this time next week.

[**After lunch treatment**]

The selected product will be given at 3 pm, in this exact place in one week from today. In case the binding product is a sandwich, we will have available fresh sandwiches made on the day of delivery. To ensure you get your pen or sandwich next week, you will be given a coupon to be redeemed, and my own professional card, in case you have any problem showing up. You can stop by my office to get the product in a weeks' time, if for some reason you are not able to pick it up from this room this time next week.

[Common text for before and after lunch treatments]

However, you will have to pay for the product today when the session is finished. The market price for the pen/sandwich determined from this auction (3rd highest price) will be

deducted from the participation fees of the highest bidders.

Are there any questions?

Please, take the pink slip number 1 and write down the maximum you are willing to pay to purchase the sandwich and the pen, respectively.

After you've finished writing your bids, the monitor will go around the room and collect the bid sheets. The monitor will then rank bids from highest to lowest, determine the 3rd highest price and the persons with bids above the 3rd highest price for each product, respectively.

The bid is private information and should not be shared with anybody else. Please be quiet while the auction is carried out.

[Once the first round is finished, second round starts]

Now, please, take the pink slip number 2 and write down the maximum you are willing to pay to purchase the sandwich and the pen, respectively.

[Same procedure is followed. Once the second round is finished, third round starts]

Now, please, take the pink slip number 3 and write down the maximum you are willing to pay to purchase the sandwich and the pen, respectively.

[Experimenter collects bid sheets]

[The experimenter asks one person to draw a number from an urn; Number determines binding round. The experimenter asks one person to draw a number from an urn. The number determines the binding product.]

[IDs of highest bidders and 3rd highest price for the binding product and round are determined and announced.]

------Page break------

Final questionnaire

The final task involves filling out a questionnaire. I will distribute the questionnaire in a minute. Please make sure your ID number is on the top left corner and raise your hands when you are done.

Thank you very much for your participation!

------Page break------

B Appendix: Additional figures

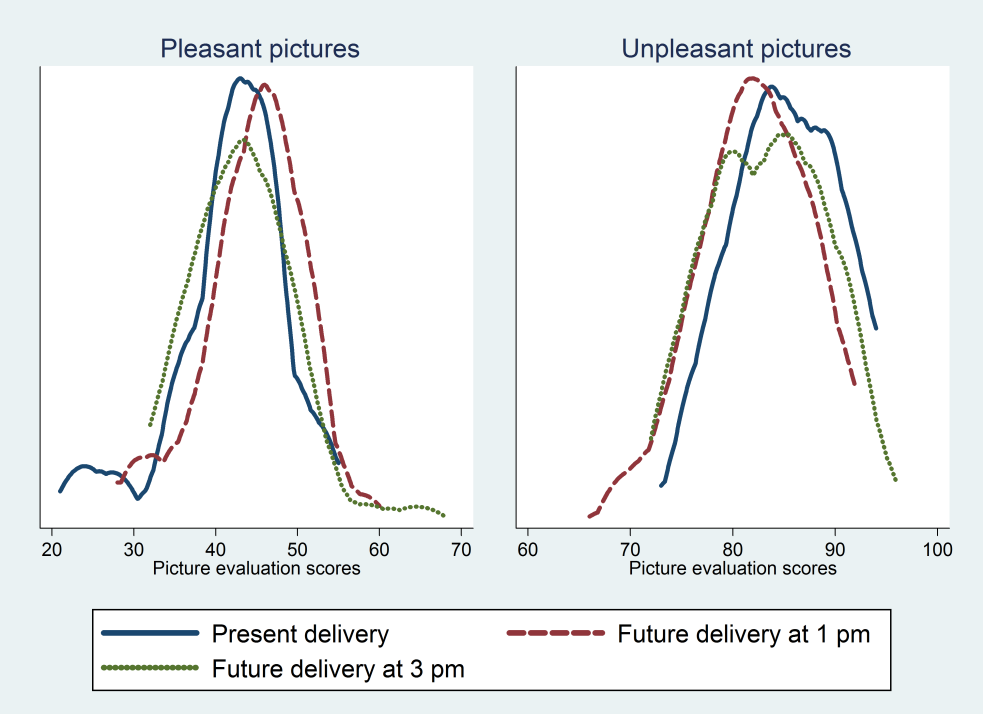


Figure B.1: Kernel density estimators of picture evaluation scores by treatment