Sorting with shame in the laboratory

Ong, David

University of California, Davis, Peking University HSBC School of Business

27 October 2008

Sorting with Shame in the Laboratory

David Ong*
Economics Department
University of California
Davis, CA 95616
dvdng@ucdavis.edu

July 29, 2009

Abstract

Trust is indispensable to fiduciary fields (e.g., credit rating), where experts exercise wide discretion on others behalf. Can the shame from scandal sort trustworthy people out of a fiduciary field? I tested for the possibility in a charitable contribution game where subjects could be "ungenerous" when unobserved. After establishing that "generosity" required a contribution of more than $6, subjects were given the choice of contributing either $5 publicly or $0-$10 privately. Almost all control subjects chose to contribute privately less than $2. The majority of treatment subjects, after being told the prediction that they were unlikely to contribute more than $2, if they contributed privately, contributed $5 publicly. This suggests that the mere belief that a subject would exploit the greater discretion and unobservability of a *fiduciary-like* position can deter entry into such a position. Thus, scandals that create such a belief could repel shame-sensitive people from that field – possibly to the detriment of the field and the economy as a whole. The shame externality of a scandals on private judgments may also been seen in politically correct speech after demonstrated racial prejudice of others.

JEL Codes: C91, C72, H41, H42

Keywords: shame, psychological game theory, beliefs preferences, charitable contributions game, fiduciary

Acknowledgement 1 *I would like to thank Giacomo Bonanno, Klaus Nehring and Burkhard Schipper for their advising. I would also like to thank Richard Robins, Guido Kuersteiner, Scott Carell, Travis Lybbert, Doug Miller, Chris Knittel, Yoonie Chung, Scott Niven, Will Ambrosini, Matthew Pearson, Santiago Buccaran, Olena Sloan, and*

*Peking University HSBC School of Business starting fall 2009

Jeffery Graham for their feedback on various parts of the writing, as well as the participants of the Applied Micro Brownbag.

1 Introduction

> President Obama called Wall Street bankers "shameful" for giving themselves nearly $20 billion in bonuses as the economy deteriorated and the government spent billions to bail out some of the nation's most prominent financial institutions. [Stolberg and Labaton, 2009]
>
> "I'd almost rather say I'm a pornographer," said a retired Wall Street executive. [Segal, 2009].

Trust is indispensable to fiduciary fields, where experts exercise wide discretion according to unobservable, subjective judgments. Clients of doctors, dentists, credit rating agencies, investment bankers, clergy, accountants...teachers need to trust the fiduciary for the same reason that they need their services – lack of expertise. An unmeasured (to my knowledge) consequence of recent scandals among financial fiduciaries is that persons most sensitive to shame might avoid scandalized tasks, leave, or never enter the profession. They may opt for non-fiduciary work where they are fully observed, and therefore, will be rewarded for moral behavior, instead of fiduciary work, where they are unobserved but suspected of immoral behavior due to the taint of scandal. If shame sensitivity positively correlates with trustworthiness, scandals could do grave damage to a profession and make future scandals even more likely by causing trustworthy people to exit and untrustworthy people to enter [1].

Though the issue of whether the shame from a scandal can sort people in fiduciary fields is an empirical question, shame aversion is not measured in job interviews. And, even if it were, since we want to measure the sorting power of shame, we would want to measure those people *who would have but did not apply* for the job. Thus, to see if scandals can sort, a controlled experiment is required.

[Tadelis, 2007] established experimentally that betrayals of trust can be deterred by the threat of *mere observation of that betrayal*: shame. However, whether the suspicion incited by *others'* shameful actions when *unobserved* could deter a person from entering into a similar

[1] Shame may have sorted the more trustworthy people out of:

1. Accounting after the indictment of Arthur Andersen.
2. Credit rating agencies after the conflict of interest scandals associated withe the internet stocks bubble.
3. Politics after a major corruption scandal.
4. The Catholic clergy after the pedophilia scandal.
5. The mortgage lending business after the recent subprime mortgage crisis.

unobserved situation has yet to be addressed. This is what is tested in the following public contribution game where shame is induced by the belief that one will be ungenerous when unobserved.

36 subjects spent about 20 minutes filling out a 50 question psychological test which they were told was to predict their likely level of generosity to a famous charity. After they revealed that 'generosity' required a contribution of more than \$6 of the \$10 they would earn, they were given the choice of contributing either \$5 publicly or \$0-\$10 privately. $\frac{9}{10}$ of control subjects contributed privately, less than \$2. $\frac{10}{26}$ of treatment subjects contributed \$5 publicly, after being told that given their low test scores, they were unlikely to contribute more than \$2, if they contributed privately. The level of significance for the treatment was 9%. The increased willingness to pay to seem generous suggests that the *mere belief* that a subject might exploit the wide discretion and unobservability (e.g., give \$0-\$10 unobserved) of a fiduciary like position can deter entry into such a position. Thus, scandals which create such beliefs could change a fiduciary field by repelling shame-sensitive people – possibly to the detriment of the field and the economy as a whole, if shame sensitivity is positively correlated with trustworthiness. This result that shame can sort people out of situations in which they might exploit moral hazard is consistent with the predictions of the pooling and separating equilibria of [Ong, 2008a][2]. To my knowledge, there are no other papers on belief or 'shame externalities'.

There are broader applications for this notion of shame externality since subjective judgments are ubiquitous, for instance, in hiring and promotion decisions by managers[3]. Scandalous prejudicial hiring practices can impose a belief externality on similar unobservable subjective judgments, which may result in public but suboptimal actions or appeasing speech acts – political correctness[4].

The psychology literature has focused upon measuring shame, but not its externalities. See [Tracy et. al, 2007] for a recent compilation of significant research in psychology.

This paper is also relevant to the debate on whether people act altruisticaly because of moral preferences, or as posited here, due to preferences over the beliefs of others. The treatment announcement in my experiment, "According to our past experience,..." was uninformative of anything other than beliefs. Subjects had been given full information about their possible choices and payoffs. Subjects who chose to contribute privately *did in fact*

[2] This confirms the predicted separating (Eq. 4) and pooling (Eq. 1) equilibria in [Ong, 2008a].

[3] "But outsiders or lower-level employees are seldom privy to the complex deliberations and the raft of subjective judgments that go into the selection of the top people in any large, complex organization." See [Loury, 1996].

[4] "...Consider "diversity training." Texaco has pledged some \$35 million for employee workshops on race relations...I doubt that anyone astute enough to rise to the top of a major corporation really believes that diversity workshops are the way to get blacks and whites to work together with mutual respect. But few will now dare give candid expression to that view. Hiring a diversity consultant is a primary way for the company to show its concern for minority sensibilities."[Loury, 1996][Eichenwald, 1996]

contribute about $1.5. If we assume that nothing but beliefs was changed with an uninformative message, then the governing norm could not have changed. However, the propensity to contribute $5 publicly did increase after the treatment.

The outline of the experiment follows. Data analysis is in section 3. The rationale of the experiment is in Section 4. Possible issues with the experiment are addressed in Section 4.5. This is followed by the conclusion and appendices containing materials used for the experiment. Appendix D contains a discussion of robustness checks.

2 Experimental Design

1. Advertisements for subjects with the heading, "Make $10 in 40 minutes," were placed around campus and on Facebook.

2. Upon arrival, I read the "Instructions and Consent" of Appendix A to the subject ("Bob" for convenience) and walked him through the experiment.

3. Bob took a standard psychological test that measures guilt and shame sensitivity (TOSCA-3), which contains 17 questions with 4 or 5 parts each that requires about 20 minutes to complete. Bob was told that the test was to predict his likely level of generosity to Doctors Without Borders (DWB) a famous charity. I added a question about Bob's major and whether he had contributed to DWB within the last year. There were no other identifiers. Bob scored his own test to maintain his anonymity.

4. Bob was then asked how much "generous" and "ungenerous" types of UC Davis students would give of the $10 that they would earn from the experiment. See Appendix B for the survey. The prior subject ("Alice" for convenience) was called in from surfing the web to witness this. (The first prior subject was a student confederate.)

5. If Bob was in the control group, he was told that a prediction based upon his test score about his likely level of contribution would not be made. If Bob was in the treatment group, he was told that a prediction would be made.

6. In the **control group**, Alice read out to Bob, "Do you want to choose the private option, where you can contribute whatever you like or contribute $5 here as you hand in the test?".

7. In the **treatment group**, before Bob was given the choice between public or private contribution in step 6, Alice asked Bob, "Is your score below 438?" If Bob said yes, Alice then read out, "According to our past experience, you are not likely to contribute more than $2, if you choose the private option." He was then given the choice in step 6.

8. Bob was paid either immediately before he was given the choice in steps 6 or 7, or immediately after, in one case deviating from the order in the Instructions and Consent. These constitute different treatments and are elaborated upon in Appendix D on Robustness.

9. Bob followed through with his choice. If Bob chose the private option, he would walk into room 109 next door, close the door and put whatever money he wanted to contribute with his test into an envelope, and then, into a sealed box. Bob had been told that the box would not be opened until at least three other subjects had done the same.

3 Results

As mentioned in part 8 directly above, each treatment involved an announcement and one of two orders of payment. The significant effect in Treatment 2, where the choice of public or private contribution occurred before payment suggests that the treatment effect in Treatment 4, where the order was reversed, is unlikely to have been due to the order of payment alone. Vice-versa for Treatment 2. See Appendix D on Robustness on these treatments for the calculation of the likelihood that each treatment effect occurred by chance. We have grounds to regard the treatment effect as being due to the announcement, made more or less effective by a particular order of payment, and to test for the statistical significance of the combined treatments.

3.0.1 Tests of Significance for the Combined Treatments

There are two possible extreme null hypotheses:

1. H_{01} =all of the subjects were rational and cared only about money.

2. H_{02} =all subjects misunderstood the game and chose independently of the treatment.

In the case of H_{01}, we do not need any formal statistics. It is rejected even if one subject chose public contribution since it is monetarily dominated. H_{02} can be ruled out with the Fisher's Exact-Boschloo test.

Fisher's Exact-Boschloo Test for All Treatments in both 2008 and 2009 The Fisher's Exact-Boschloo test is used when the scores from two independent random samples: here treatment and control, all fall into one or the other of two mutually exclusive classes: here public and private. See [Schlag, 2008] for more details on the Fisher's Exact-Boschloo test. Let N=number of subjects in the treatment.

N = 36	**Public**	**Private**
Combined Treatments	10	16
Control	1	9

(Table 1)

The unconditional p-value[5] or "the maximun probability under H_{01} of observing the test statistic equal to or more extreme than the value observed in the data"[Berger, 1996] for the two-tailed test is 0.093[6]. The case where only *some* of the subjects misunderstood the game and chose public by accident (and therefore some other subjects chose public due the treatment effect), cannot be less significant than 9%. Also, this level of significance for all treatments has to be regarded as a kind of lower bound on significance for the most effective treatment.

4 Rationale for Experimental Design

4.1 Increasing Subjects Shame Sensitivity

TOSCA-3 asks subjects to imagine themselves in 17 scenarios in which they might feel shame. I used this test to prime subjects for the possibility of shame, because in effect, it asks subjects to practice feeling ashamed in imagination. An example of a question from TOSCA-3:

1. You make plans to meet a friend for lunch. At 5 o'clock, you realize you stood him up (you didn't show up when you said you would).

a) You would think: "I'm inconsiderate." 1---2---3---4---5 not likely very likely

b) You would think: "Well, they'll understand." 1---2---3---4---5 not likely very likely

c) You'd think you should make it up to him as soon as possible. 1---2---3---4---5 not likely very likely

d) You would think: "My boss distracted me just before lunch." 1---2---3---4---5 not likely very likely

Figure 1: TOSCA-3 questions.

According to the psychology literature, shame is due to beliefs about others beliefs [Tracy et. al, 2007] that one has violated some norm or standard of behavior. Whatever shame Bob might feel from taking the private option after Alice announces her belief that he will act ungenerously, I tried to leveraged that shame further by the apparent scientific validity of that belief.

[5] See [Greenland, 1991] for the justification for the use of the unconditional p-value.

[6] I would like to thank Karl Schlag for making me aware of this test. See [Schlag, 2008] for his notes. See [Berger, 2005] for the calculation software.

Bob scored his own test to preserve his anonymity. The score was a weighted average of test answers based on the hypothesis that generosity is correlated with guilt sensitivity. The score was heavily weighted by the answer of an added question – whether the subject contributed to Doctors Without Borders (DWB) in the last year. The score was designed to camouflage the relationship between the numerical values of the answers and our prediction for the subjects level of contribution, so as to make it less likely that the subject would try to game the test (e.g., answer yes to the DWB question and be confident that we believed that he would contribute generously when observed) and hence obviate the need to prove his generosity by giving $5 publicly.

4.2 Establishing Norms of Generosity

On average, subjects estimated that the generous type would contribute more than $6 and the ungenerous less than $1. See Table 1-6 below and Appendix C for the data. Bob's estimate was intended to credibly establish the type space: 'generous' and 'ungenerous', with respect to which Bob could signal his own type (e.g., contribute more than the ungenerous type so as to decrease the probability of being thought ungenerous). The accuracy of the prediction did not matter for the experiment. What mattered was that Bob credibly committed himself to a high and therefore costly (above $2) standard of generosity in front of Alice and the experimenter. In fact, Alice, who may take a low estimate personally, was there in part to bias Bob's estimate upwards.

4.3 The Choice Between Observable (Public) and Unobservable (Private) Contributions

The public option of contributing $5 was restricted. Therefore, it was (monetarily) dominated by the private option, where the subject could contribute $0-$10. However, unlike the private option, it permitted the subject to make evident to observers that he was not the 'ungenerous' type. Hence, it may not be dominated if non-monetary payoffs are taken into account.

4.4 Treatment

Alice only asked, "Is your score below 438?" instead of the actual score because that could be used to identify Bob with his contribution, via his test which he put in the same envelope, thus undermining the unobservedness of the private option. His score could only be above that number if he contributed to DWB within the previous year. If he answered "yes", Alice read out to him, "According to our past experience, you are not likely to contribute more than $2, if you choose the private option[7]." This announcement of the expectation of low

[7]This estimate was gleened from past pilot experiments with other designs.

contribution levels was designed to induce shame – conditional on the private option being taken. It played the same role as the drug firm representative's remark "One hand was the other" in "Fishy Gifts" [Ong, 2008a], where reciprocation was shameful.

As mentioned in the experimental design, I varied the order of payment. See Appendix D for the discussion.

4.5 Possible Problems with the Experiment

A number of the possible problems with this attempt to capture a shame spillover in the laboratory can be ruled out by the data.

1. If subjects did not regard the private option as actually unobserved, then, contrary to my actual results, there should not have been any significant treatment effect on the probability with which subjects chose the private option.

2. If subjects did not think that their contribution would actually go to DWB, then contrary to my results, they would only have chosen the private option and contributed nothing.

3. There could have been shame in the private option even when subjects were untreated. This shame would not explain the change in behavior when subjects were treated.

4. The public contribution could be due to a self-image preference. However, the subjects in the control group, those who were not treated with the announcement of observers' beliefs, did not mind taking the private option and making a low contribution. *Presumably*, subjects self image is independent of an announcement of observers' beliefs about what a subject will do when unobserved.

5. There is the possibility that Bob's choice was not entirely independent of Alice's since Alice read out the prediction to Bob. However, the monetary payoffs of Bob's choices were fully revealed. Therefore, Alice could only have communicated her beliefs about Bob in her announcement. The effect of this belief, i.e., shame, was what was being tested for. For that purpose, it didn't matter that Alice had been a prior subject. Thus, though the prior subject read out the prediction for the current subject, the independence assumption necessary for the Fisher's Exact-Boschloo test still applies.

6. Subjects could have also guessed the motive of the experiment – to induce shame that required a costly action of the public contribution of $5 to avoid. In that case, they wouldn't have contributed anything and their TOSCA score would have been unusual, contrary to my data.

7. One could argue that the treatment effect was due to experimenter demand. However, the announcement was only about the beliefs of the experimenter and not about the experimenters preferences. The subject could have been responding to his preferences over the beliefs of the experimenter, but that is what is being tested for.

8. Some of the instructions were unclear. For example, the subject was not told what would happen if there were not 3 other subjects who made private contributions. No one asked and I did not explain how or why a psychological test would be used to predict a subject's level of generosity. I did not try to dispel these ambiguities due to time or budget constraints or because I didn't want the subject to think too much about the experiment. In any case, confusion should lead to greater randomness in lower significance levels.

5 Conclusion

The results of the above experiments suggest that the shame spillovers from scandals can sort people out of fiduciary-like positions. The correlation between shame and guilt as measured by TOSCA-3 is estimated at 0.5 [Tangney and Dearing, 2002]. If people were being sorted by shame sensitivity, they would also be sorted out by guilt sensitivity. Experimental results from trust games like [Charness and Dufwenberg, 2006] suggested that some notion of "guilt" increases reciprocation of trust. Together, they suggest that those most likely to be trustworthy are also most likely to be sorted out of fiduciary position by a scandal. Then, not only would scandals damage the field, the damage to the reputation of the field would select for people who would further damage the field. President Obama's shaming of Wall Street employees [Stolberg and Labaton, 2009][Segal, 2009] could therefore have exactly the opposite effect from what he intended.

But, even without a scandal, fiduciary positions should attract the least trustworthy people because they have the most to gain or least to lose from betraying trust. According to Raymond W. McDaniel of Moody's[McDonald, 2008]:

> "The real problem is not that the market ... underweight[s] ratings quality but rather that in some sectors, it actually penalizes quality. ... It turns out that ratings quality has surprisingly few friends: issuers want high ratings; investors don't want ratings downgrades; short-sighted bankers labor short-sightedly to game the ratings agencies." McDaniel then tells his board: "Unchecked, competition on this basis can place the entire financial system at risk." Furthermore, though Moody's has "erected safeguards to keep teams from too easily solving the market share problem by lowering standards. This does NOT solve the problem."

Given this problem of adverse selection into fiduciary professions, how is it possible that fiduciary professions function at all? What institutional measures exist to counteract the adverse selection to fiduciary fields? In [Ong 2008b], I model how institutional arrangements in fiduciary professions, like pro-bono work, can save the reputation of a field by sorting people who might exploit trust out of the field.

5.1 Appendix A: Instructions and Consent

This experiment will proceed as follows:

1. You will be asked to take a standard psychological test of 17 questions that we will use to estimate your likely level of generosity to Doctors Without Borders (DWB), an organization which brings western doctors to parts of the world where medical care is urgently needed but not available.

2. To preserve anonymity, you will score your own test using an Excel spreadsheet. Write down your score on the piece of paper provided, but do not show it to us. Then close the spreadsheet without saving.

3. Before another UCD student, you will be asked to state an estimate of how much,

a. a generous type of UCD student would give of the $10 that they earn to DWB.

b. an ungenerous type of UCD student would give of the $10 that they earn to DWB.

4. After you make your estimate, you will be paid $10 and asked to sign for it. *After you sign for it, the money is yours.*

5. Then, you will be given the opportunity to donate $5 when you hand in the test, or any amount you think appropriate anonymously in room 109. If you take the anonymous option, please put the test and the money in the envelope provided. A receipt from Doctors Without Borders for the cumulative amount of money will be posted on the web at the end of the experiment in a few weeks.

6. Before you contribute, we may or may not score your test and inform you of how much you are likely to contribute should you choose the anonymous option. If we score your test, the previous participant will read you the prediction.

7. **This test is anonymous.** There is nothing to identify you with your contribution or your test score. For the purpose of the experiment, we will only record your major. For the purpose of paying you, we will keep a receipt of your guess and the fact that we paid you. You will be asked to stay until the next participant makes their choice. That way, you can also be sure that the box remains unopened, thus preserving your anonymity. We would not open the box until at least 3 participants have taken the anonymous option.

I understand these instructions and would like to participate in the experiment

Name_____

Signature_____ Date_____

5.2 Appendix B: Experimental Subject's Predictions

Circle your estimate of the average contribution of generous people.

$0
$1
$2
$3
$4
$5
$6
$7
$8
$9
$10

Circle your estimate of the average contribution of not generous people.

5.3 Appendix C: Data

Subject	**1**	**2**	**3**	**4**	**5**	**6**
Religion	Christian	Buddhist	Atheist	None	None	Catholic
Estimate: Generous	5	5	1	5	10	5
Estimate: Not Generous	0	0	0	0	2	0
Guilt Sensitivity	66	72	60	62	59	54
Shame Sensitivity	44	49	45	48	54	25
Informed	Yes	Yes	Yes	Yes	No	Yes
Public	No	No	No	No	No	No
Contribution	6	0	10	5	4	3

(Table 4: Data from Treatment 1 (August 2008).)

Subject	**1**	**2**	**3**	**4**	**5**
Estimate: Generous	10	10	5	5	5
Estimate: Not Generous	3	0	2	0	0
Guilt Sensitivity	52	57	68	64	59
Shame Sensitivity	54	49	42	57	46
Informed	Yes	No	Yes	Yes	Yes
Public	No	No	No	No	No
Contribution	3	3	0	3	0

(Table 5: Data from Treatment 3 (August 2008))

Subject	**1**	**2**	**3**	**4**	**5**
Religion	None				
Estimate: Generous	6	5	5	5	6
Estimate: Not Generous	0	0	2	0	0
Guilt Sensitivity	54	64	56	50	67
Shame Sensitivity	41	37	41	27	41
Informed	Yes	Yes	Yes	Yes	Yes
Public	Yes	No	Yes	No	Yes
Contribution	5	1	5	2	5

(Table 6: Data from Treatment 4 (August 2008))

Subject	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Estimate: Generous	7	5	10	10	3	6	5	5	6
Estimate: Not Generous	0	0	0	2	0	0	0	0	1
Guilt Sensitivity	67	59	67	71	66	68	70	56	47
Shame Sensitivity	49	28	53	46	52	57	55	43	47
Treated	Y	Y	Y	Y	Y	Y	Y	Y	Y
Public	N	N	N	Y	N	N	N	N	N
Contribution	5	0	0	5	0	3	0	X	5

(Table 6: Data From Treatment 2 (April 2009))

Subject	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**Av**
Estimate: Generous	10	5	7	3	7	10	8	3	**6.5**
Estimate: Not Generous	0	0	2	0	0	3	1	0	**0.5**
Guilt Sensitivity	45	60	61	45	67	58	67	61	**61**
Shame Sensitivity	32	43	45	33	55	47	46	48	**46**
Treated	N	N	N	N	Y	Y	Y	Y	
Public	N	N	Y	N	Y	Y	N	N	
Contribution	0	2	5	0	5	5	X	X	**2.1**

(Table 7: Data From Treatment 2 (April 2009))

6 Appendix D: Robustness: Variations in the Treatment

6.0.1 Treatment 1: TOSCA-3 with Subject Identified by Religion

My original modified TOSCA-3 included an identifier for the religion of the subject. Average contributions levels jumped from \$2 in pilot experiments[8] to \$5. There was also significant negative correlation between guilt and contribution levels. See first four columns of Table 4 in Appendix C. These unexpected results would still be predicted by a treatment of my model in [Ong, 2008a], developed further in [Ong 2008b]. Some subjects who belonged to religions that were not known for generosity may have wanted to contribute more because they wanted to make their religion look better to the experimenter or the audience of the paper. For example the atheist contributed \$10 and wrote that the religion was "strongly atheist". Others may be content to free-ride on the good reputation of their religion (e.g., the Buddhist gave \$0). The negative correlation between the guilt of the subject and contribution level would make sense if the low guilt people sorted into less charitable religions and hence, were more burdened with shame. However, I wanted to focus on sorting by shame alone. I decided to drop the religion question and restart. After getting the major result of the experiment discussed above in Table 2 below, I decided to do further experiments with surveys that included the religion question to see if I could show that it indeed made a significant difference to the level of contributions. But, upon subsequent trials, the contribution level went down and the negative correlation disappeared. See columns 5-7 of Table 4.

6.0.2 Treatment 2: Surprise Reversal of Order

When I restarted the experiment, my first Alice read out the choice between public or private contribution before I could pay my first Bob, contrary to the lines 4 and 5 of the instructions (See Appendix A.). There was a noticeable change in Bob's level of agitation at the choice of public or private. I took the fact that he also chose the public option as significant. I decided to keep that order reversal for all 7 subsequent subjects that week, both treated and untreated, so information was still the only thing that varied across subjects. This is the data in table 2 below.

[8] I made some major changes to the experiment after presenting the results of a set of pilot experiments at a departmental brownbag. Among them was the inclusion of a question about the subject's religion in the modified TOSCA-3.

Treatment 2 (September 2008)									**Average**
Subject	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	
Estimate: Generous	10	2	7	9	10	10	9	3	**7.5**
Estimate: Not Generous	0	0	0	0	0	0	0	0	0
Guilt Sensitivity	64	63	65	58	72	48	51	70	**61**
Shame Sensitivity	47	58	52	59	66	45	48	57	**54**
Treated	*Y*	*Y*	*Y*	*Y*	*N*	*N*	*N*	*N*	
Public	*Y*	*Y*	*Y*	*Y*	*N*	*N*	*N*	*N*	
Contribution	5	5	5	5	2	2	1	1	**3.3**

(Table 2)

All 4 subjects of the treatment group chose the public option of contributing \$5. All 4 subjects of the control group chose the private option and gave less than \$2.

Fisher's Exact-Boschloo Test for Treatment 2 (August 2008 Rounds)

N = 8	**Public**	**Private**
Treatment	4	0
Control	0	4

(Table 3)

The p-value was 2.8%. Note that this result is not categorically better than Treatment 4, where there was no change in order, but 3/5 subjects chose the monetarily dominated option of giving in public. The level of significance there was 16%.

Several reasons come to mind as to why I could now get 100% treatment from the change in order between lines 4 and 5:

1. The change in order diminished the endowment effect, which would have made the subject want to keep more of the money, countering the effect of shame, which would have made the subject want to seem generous by giving away more money.

2. The strong effect was due to the change in order only but not the surprise. Under this hypothesis, the payment would not intervene between the admission of the subject that generosity required contributions greater than \$5 and the announcement that the subject was not likely to give more than \$2 The lack of effect in Treatment 3, where there was a change in order but no surprise, suggests the unlikeliness of this possibility.

3. The surprise change in order could have intensified the effect of shame. If so, this may be why I was able to measure the effect of shame even though there was only two observers (the experimenter and Alice). [Tadelis, 2007] used at least 10 observers to get his effect.

4. The deviation could have suggested the possibility of further deviations in the subject's mind, e.g., that he risked not being paid at all if he chose the private option. In the advertisements for subject, they were promised $10 in 40 minutes for taking a survey and splitting some money. They also signed a consent which explicitly said that they would be paid. In this context, non-payment would be such a breach of professionalism that it's hard to imagine that any subject could entertain that possibility.

6.0.3 Comparing Treatment 2 in 2009 to 2008

I did 14 more subjects in 2009 (See Table 6 and 7 in Appendix C for the data.) to try to corroborate Treatment 2 (for a total of 36 subjects[9] for all treatments for all years). There was a decrease of $1.2 for Treatment 2 from $3.3 (2008) to $2.1 (2009) in the average contribution level of subject. Average shame sensitivity, as measured by TOSCA 3, went down by 15%, which suggests that the subjects could have been different. The decrease in contribution level would be predicted if belief preferences were contingent on wealth and wealth was expected to be lower due to the intervening financial crisis. Nonetheless, the level of significance was still greater than the other treatments, excluding Treatment 2.

6.0.4 Treatment 3: Reversal of Order But No Surprise

To test whether the reversal of order between line 4 and 5 alone in Appendix A, as opposed to the reversal order *and* the surprise deviation from instructions was responsible for the perfect correlation, I removed line 4, "After you make your estimate, you will be paid $10 and asked to sign for it. After you sign for it, the money is yours." from the 'Instructions and Consent' form in Appendix A and merely paid subjects after they made their choice. Then, 0/4 subjects who were in the treatment group chose the public option, though one subject took about 20 seconds to make his choice. See Table 5 in Appendix C. Thus, it became less plausible that reversal of order alone increased the probability of the choice of the public option.

6.0.5 Treatment 4: No Reversal of Order

The unimportance of the mere reversal of order of lines 4 and 5 in Treatment 3 without surprise and that in Treatment 2 with surprise was further confirmed when I re-established the original order. Then, 3/5 of subjects chose the public option. See Table 6 in Appendix C. The increase in those who chose public suggests that the strong treatment effect in Treatment 2 was due either to the surprise or to both the surprise and reversal of order. More data is necessary to make an firm conclusion.

[9] I lost access to the rooms for this experiment after these subjects and was only able to regain it in spring 2009.

N = 9	**Public**	**Private**
Treatment	3	2
Control	0	4

As mentioned above, the p-value for this treatment was 16.7%.

6.0.6 All Treatments from 2008

The financial crisis occurred between my rounds of experiments. The ensuing changes in average wealth could make it less shameful to be ungenerous. Thus, it may be good to separate the data from before the crisis to after.

If all of 22 subjects of the treatments in 2008 were pooled and called one experiment, the result would be close to being significant at the 10% level for the two tail test[10]:

$N = 22$	**Public**	**Private**
Treatment	7	9
Control	0	6

The p-value was 12% for the two tailed test[**?**].

6.0.7 Excluded Data from 2009

The structure of the lab is not ideally suited for this experiment. In the 2009 experiments, Subject 8 saw 7's choice of public or private. Subject 17's arrival interrupted 16's choice of public or private. Subject 17 placed his score in full view after being told that he should hide it, perhaps in an effort to inform observers of his level of generosity, without having to signal with the costly choice of the $5 public contribution. Xs in the contribution box of Table 9 stand for "excluded" from the analysis.

[10] The total of all treatments in Table 1 was calclated by adding the 14 subjects from 2009 to the 22 subjects from 2008.

$N = 37$	**Public**	**Private**
Treatment	$7 + 3$	$9 + 7$
Control	$0 + 1$	$6 + 3$

References

[Battigalli and Dufwenberg, 2007] Battigalli, P, Dufwenberg, M, "Guilt in Games," American Economic Review, Volume (Year): 97, 2007

[Battigalli and Dufwenberg, 2008] Battigalli, P, Dufwenberg, M, "Dynamic Psychological Games," Journal of Economic Theory, Elsevier, vol. 144(1), pages 1-35, January

[Berger, 1996] Berger, RL, "More Powerful Tests from Confidence Interval P-Value", The American Statician, Vol 50, No. 4, P. 314-318, Nov. 1996

[Berger, 2005] Berger, RL, "Exact Unconditional Homogeneity/Independence Tests for 2X2 Tables", http://www.stat.ncsu.edu/exact/

[Benedict, 1946] Benedict, R, The Chrysanthemum and the Sword. Boston: Houghton Mifflin, 1946

[Charness and Dufwenberg, 2006] Charness, G, Dufwenberg, M, "Promises and Partnership," Econometrica, Econometric Society, Volume: 74, Pages: 1579-1601, Issue: 6 November, 2006

[Dufwenberg and Gneezy, 2000] Dufwenberg, M, Gneezy, U, "Measuring Beliefs in an Experimental Lost Wallet Game," Games and Economic Behavior 30, 163–182, 2000

[Eichenwald, 1996] Eichenwald, K, "Investigation Finds No Evidence of Slur On Texaco Tapes", The New York Times, November 11, 1996

[Gehm and Schere, 1988] Gehm, TL, Scherer, KR, "Relating situation evaluation to emotion differentiation: Nonmetric analysis of cross-cultural questionnaire data," in K. R. Scherer (Ed.), Facets of emotion: Recent research (pp. 61-77) Hillsdale, NJ: Lawrence Earlbaum, 1988

[Greenland, 1991] Greenland, S, "On the Logical Justification of Conditional Tests for Two-by-Two Contingency Tables," The American Statistician, Vol. 45, 1991

[Hurely and Shogren, 2005] Hurely, TM, Shogren, JF, "An Experimental Comparison of Induced and Elicited Beliefs" The Journal of Risk and Uncertainty, 30:2; 169–188, 2005

[Krishnamoorthy, 2006] Krishnamoorthy, K., Handbook of Statistical Distributions with Applications, Chapman & Hall/CRC Press, 2006

[Lewis 1971] Lewis, HB,Shame and Guilt in Neurosis. New York: International Universities Press, 1971

[Loury, 1996] Loury, Glenn, "The Racism We Condemn", New York Times, November 24, 1996

[McDonald, 2008] MacDonald, E., "The Credit Rating Agencies' Moment of Shame", Fox Business, October 23, 2008. http://emac.blogs.foxbusiness.com/2008/10/23/the-credit-rating-agencies-moment-of-shame/

[Nyarko and Schotter, 2002] Nyarko,Y and Schotter,A, "An Experimental Study of Belief Learning Using Elicited Beliefs," Econometrica 70, 971–1005, 2002

[Ong, 2008a] Ong, D, "Fishy Gifts: Bribing with Shame and Guilt," Mimeo, 2008

[Ong 2008b] Ong, D, "Pro-Bono Work and Trust in Fiduciary Professions," Mimeo, 2008 (available upon request)

[Pierce, 1970] Piece, A, Fundamentals of Non-parametric Statistics, Dickenson Publishing Company, Belmont, California, 1970

[Preacher and Briggs, 2001] Preacher, KJ, & Briggs, NE, Calculation for Fisher's Exact Test: An interactive calculation tool for Fisher's exact probability test for 2 x 2 tables [Computer software]. Available from http://www.quantpsy.org, 2001

[Ritholtz, 2008] Ritholtz, B, "S&P: We Knew Nothing! Nothing!" The Big Picture,Wednesday, October 22, 2008 http://bigpicture.typepad.com/comments/2008/10/sp-its-not-our.html

[Schlag, 2008] Schlag, K, "Exact Hypothesis Testing without Assumptions - New and Old Results not only for Experimental Game Theory", July 5, 2008

[Segal, 2009] Segal, D, "Wall St., a Financial Epithet, Stirs Outrage," New York Times, February 2, 2009

[Siegel and Castallan, 1998] Sidney, S. Castallan, J., Nonparametric Statistics for The Behavioral Sciences, McGraw-Hill Humanities; 2 edition, January 1, 1988

[Sorkin, 2008] Sorkin, AR, "Rating Agencies Draw Fire on Capitol Hill DealBook," New York Times, October 22, 2008

[Stolberg and Labaton, 2009] Stolberg, SG, Labaton, S, "Obama Calls Wall Street Bonuses 'Shameful'," New York Times, January 29, 2009

[Tadelis, 2007] Tadelis, S, "The Power of Shame and the Rationality of Trust", Working Paper, UC Berkeley, Haas School of Business, August 10, 2007

[Tangney and Dearing, 2002] Tangney, JP, Dearing, RL, Shame and Guilt, The Guilford Press, 2002

[Tracy et. al, 2007] Tracey, JL, Richard, RW, Tangney, JP, Campos, JJ, The Self-conscious Emotions, Guilford Press, 2007