Spring 2021

Financial Decision Making Under Stress

Narek Vartan Bejanyan
Claremont Graduate University

Financial Decision Making Under Stress

By

Narek Vartan Bejanyan

Claremont Graduate University

2021

Approval of the Dissertation Committee

This dissertation has been duly read, reviewed, and critiqued by the Committee listed below, which hereby approves the manuscript of Narek Vartan Bejanyan as fulfilling the scope and quality requirements for meriting the degree of Doctor of Philosophy in Economics, concentration in Behavioral and Neuroeconomics.

Paul J. Zak, Chair

Department of Economics

Claremont Graduate University

Professor of Economic Sciences, Psychology & Management

Director, Center for Neuroeconomics Studies

Joshua Tasoff

Claremont Graduate University

Associate Professor of Economic Sciences

Tom Kniesner

Claremont Graduate University

University Professor

Abstract

Financial Decision Making Under Stress

By Narek Vartan Bejanyan

Claremont Graduate University: 2021

The aim of this paper is to examine the link between individual's financial decisions and stress. In a laboratory experiment, Holt and Laury lottery was used to elicit participants risk preferences. The Cold Pressor Test was used to induce a safe level of stress. In the treatment group, 62% of participants are risk averse; versus 76% in the control group. Additionally, inconsistent financial decision making was observed in both groups: 55% of participants in the treatment group versus 30% in the control.

Table of Contents

Chapter 1

Financial Decision Making Under Stress

Introduction

Decision making under stress has been a determining factor for individuals in society that dictates whether they have a successful decision making outcome or not. From an evolutionary perspective, an individual's decision making abilities under stress, for example when a lion attacks a person's camp, determines their chances of survival. Society has evolved to the point where we do not often face life threatening situations like a lion attack. However, there are still instances where we commonly find ourselves under stress: taking exams, buying stocks, running in/out of a burning building, and making financial decisions, to name a few.

Therefore, it is of great relevance to examine financial decision making under stress. Financial decision making has been studied in many different contexts in attempt to understand and increase individuals' utility. This paper presents results from a laboratory experiment which elicited participant's risk preferences under stress. The experiment tested the hypothesis of whether participants under stress exhibit risk averse, loving, or neutral behavior. The main hypothesis was tested in a laboratory experiment using a Holt Laury's lottery task to elicit participants risk preferences. The Cold Pressor Test (CPT) was used to induce a safe level of stress.

The main hypothesis is that participants under stress will exhibit risk aversion behavior when making financial decisions. Participants are less likely to take a risk in the presence of stress.

Literature Review

2.1 Holt and Laury

Holt and Laury (HL) in 2002 introduced a menu of choices tasks that can be used to estimate the degree of risk aversion as well as specific functional form. Participants in an HL designed experiment choose between two options ten times as the expected payoffs change (Holt and Laury, 2002, p.1644). Depending on when participants switch options, their risk preference is revealed (Holt and Laury, 2002, p.1646).

For the current experiment, the payoffs for the less risky option is \$4.00 or \$3.50 (option A), which is much less than the potential payoffs of \$7.00 or \$1.00 in the more risky option (option B) (Table 1). The two options are randomly presented to participants ten times. Each time the options are presented, the probability of earning a higher payoff increases and the probability of earning a lower payoff decreases. For example, in the first choice, the probability of the high payoff (\$4.00 or \$7.00) for each option is 1/10 versus for the lower payoff (\$3.50 or \$1.00) the probability is 9/10; hence, only an extreme risk seeker would choose a more risky option (Option B) (Table 1). On the other hand, even the most risk-averse person should switch over by decision 10 in the bottom row, since the more risky option (option B) yields a sure payoff of \$7.00 (Table 1).

The switching point is determined where expected payoff differences changes from positive to negative. Specifically, the expected payoff incentive to choose less risky Option A ranges from \$1.95 to -\$3.00 (Table 1). The payoffs for the lottery choices in the experiment were selected so that the switching point would provide an interval estimate of a participant's coefficient of relative risk aversion (Table 2). Furthermore, the payoff numbers for the lotteries so that the risk-neutral choice pattern of four safe choices followed by six risky choices was optimal for constant relative risk aversion in the interval (-.24,.19) (Table 2).

In literature, constant relative risk aversion is generally assumed due to functional form being logarithmic linear and computationally convenient (Holt and Laury, 2002, p.1646). Therefore, calculated risk aversion can be interpreted as: risk preference for $r < 0$, risk neutrality for $r = 0$, and risk aversion for $r > 0$ (Holt and Laury, 2002, p.1646). However, constant relative risk aversion is not a necessary assumption for an HL designed experiment.

2.2 Cold Pressor Test

Hines and Brown (1936) originally developed the Cold Pressor Test (CPT) as a procedure which is carried out by immersing an extremity in ice water (p.1). That is, one hand was immersed above the wrist in ice water ranging from 4^{O} to 5^{O} C for about 20 to 30 secs (Hines & Brown, 1936, p.2). In recent literature, experimentalists have used two components of electrodermal activity (EDA) to record and analyze participants' responses to the CPT: skin conductance level (SCL) and non-specific skin conductance responses (NS.SCRs). For example, the CPT was used in an experiment in which SCLs were utilized to develop a relative intra-individual comparison (Horstick et al., 2018, p.2). Specifically, authors demonstrated that the SCL can be used as an index of sympathetic activation as result of the effects of physical stress situations (Horstick et al., 2018, p.8). Also, Posada-Quintero et al. (2016) use EDA differences between baseline conditions and three treatments to elicit sympathetic activation: postural stimulation, CPT, and the Stroop test (p.3125). The authors found significant differences in NS.SCRs and SCL when comparing baseline to the CPT (Posada-Quintero et al., 2016, p.3125).

2.3.1 Risk Aversion: Humans

Ruixun et al. (2014) specified several measures of risk aversion that have been developed including curvature of utility functions, labor supply behavior, option prices, and others (p.17777). Risk aversion as a basic insight into human behavior has been studied academically dating back to the St. Petersburg Paradox (Ruixun et al., 2014, p17779). The paradox originated from Daniel Bernoulli's presentation of the problem and solution published in 1738 (Ruixun et al., 2014, p17779). Bernoulli's pioneering work on gambling introduced a formal framework to investigate risk aversion (Ruixun et al., 2014, p17779). That is, expected values are measured by multiplying each possible gain by the number of ways in which it can occur and then dividing the sum of these products by the total number of possible cases (Bernoulli, 1954, p. 24). There is no reason to assume that any two individuals encounter identical risks; and neither should one expect to have his desires more closely fulfilled (Bernoulli, 1954, p. 24). Similarly, not all characteristics of an individual should to be considered, only ones that pertain to the terms of risk (Bernoulli, 1954, p. 24).

Centuries after Bernoulli, in 1944 Neumann V. J. and Oskar M. developed the theory of expected utility—an individual may determine the maximum utility which can be obtained with quantities of goods at his disposal (p.30). Furthermore, given that the assumptions above the maximum is a well-defined quantity; the increase of any definite good is well-defined when added to the stock of all goods in the possession of the individual (Neumann V. J. and Oskar M., 1944, p.31).

These authors are describing the classical notion of the marginal utility of the commodity in question, or more precisely, indirectly dependent expected utility (Neumann V. J. and Oskar M., 1944, p.31). These quantities are clearly of decisive importance in the Robinson Crusoe economy (Neumann V. J. and Oskar M., 1944, p.31). If an individual is behaving rationally, then marginal utility corresponds to the maximum effort an individual is willing to exert to obtain one more unit of that commodity (Neumann V. J. and Oskar M., 1944, p.31). However, it is not clear what significance marginal utility has in determining the behavior of a participant in a social exchange economy.

A few decades after Von Neumann and Morgenstern work in the 70's, Kahneman and Tversky's groundbreaking prospect theory introduced behavioral decision making under uncertainty. Choices among risky prospects exhibit several pervasive effects that are inconsistent with the basic tenets of utility theory (Kahneman & Tversky, 1979, p.265). The certainty effect describes individuals with the tendency to underweight outcomes that were merely probable in comparison with outcomes that are obtained with certainty (Kahneman & Tversky, 1979, p.265). The certainty effect contributes to risk aversion in choices involving sure gains and contributes to risk seeking in choices involving sure losses (Kahneman & Tversky, 1979, p.265).

Second, the isolation effect describes individuals with the tendency to discard components that are shared by all prospects under consideration (Kahneman & Tversky, 1979, p.271). The isolation effect leads to inconsistent preferences when the same choice is presented in different forms (Kahneman & Tversky, 1979, p.271).

2.3.2 Risk Aversion: Animals

Stress has been studied using human subjects in various fields such as Economics, Psychology, Sociology, and Zoology. Specifically, there have been studies using animals to examine the effects of stress on decision making processes in field of Zoology. In animal literature, the energy budget rule describes risk sensitivity as the response of organisms whose goal is the maximization of Darwinian fitness in stochastic environments (Weber et al., 2004, p.430).

The energy budget rule explains animals' risk preferences like prospect theory does for human beings (Weber et al., 2004, p.434). The energy budget rule predicts risk aversion when animals are not in danger of starvation or domain of gains versus risk seeking when there is danger of starvation or domain of losses (Weber et al., 2004, p.434). Specifically, ratio comparisons are not restricted to just human comparisons of money savings (Weber et al., 2004, p.434). For example, a large amount of evidence suggests that rats use comparisons of numericities involving ratio operations (Gallistel and Gelman, 1992, p.44).

Risk aversion as a determining factor in financial decision making has been studied vigorously within previous literature. Furthermore, both human and animal experiments have studied the decision making process under optimal neural capacity. However, there is a gap in literature that is prevalent when considering individuals financial decision making under stress. This research is contributing to literature by filling in this gap with results from a laboratory experiment where participants' financial decisions were studied under stress.

Methods

3.1 Sample

Potential participants were recruited utilizing an approved email distribution list from the Center for Neuroeconomics Studies (CNS) website. In total, 68 individuals were recruited to participate in the experiment titled Risk Experiment Under Stress. The laboratory experiment was approved by the Institutional Review Board (#2914). Participants were notified before arriving that the opportunity involved completing surveys, some of which would be done after immersing their dominant hand in cold or warm water. They were also informed that immersing their hand would increase their heart rate and cause them varying levels of pain.

3.2 Procedure

Participants were asked to arrive at the Center for Neuroeconomics Studies (CNS) lab located in Claremont Graduate University, Claremont, CA. Upon arrival, participants were given a consent form with a detailed explanation of the experiment. Within the consent form, a clear description of the earning was disclosed: "You will earn a maximum of $7 on the lottery task that you will complete after you immerse your hand in ice or warm water. You will earn $.50 for every 15 secs of you holding your hand under ice or warm water, maximum amount of time allowed is 2 minutes. Possible earnings total $11" (Consent Form).

Once written consent was provided, participants received an identity masking code and were randomly assigned to control or treatment groups. Participants were informed their participation would take approximately 25-30 min. Immediately, a lab administrator screened for conflicts with increased heart rate due to the cold pressor test. Participants were excluded if they had a history of cardiovascular disorder, fainting, seizures, Raynaud's phenomenon, frostbite, and open cut/fracture on their dominant hand. Also, participants under the age of 18 were excluded.

Verbal instruction was given before beginning the experiment to assure participants understood their earnings and the CPT procedure. Participants were informed that they may remove their hand from the water when they wished. Furthermore, they should keep in mind the longer they kept their hand in the water the higher their pay will be. Participants were informed that a maximum of two minutes was allowed and the experimenter would stop the CPT only when they have reached the maximum of two minutes. Participants were not informed of how long they had their hand in water, only when they reached the maximum amount of time allowed.

The only difference between the control and treatment groups was the water temperature. The temperature of the water was recorded a few minutes prior to participants scheduled time slot. However, the temperature was not disclosed to the participants. Hence, participants had no prior knowledge if they were in the control versus treatment groups before the beginning of the experiment. Warm water temperature for the control group ranged between 20.3^{O} and 20.5^{O} C; which was approximately room temperature. Conversely, the cold water ranged between -2^{O} and $-.6^{O}$ C.

Thereafter, the experimenter prepared and attached electrodermal activity (EDA) sensors onto participants non-dominate hand. Once EDA sensors were placed, participants were asked to sit and rest for three minutes while baseline levels were recorded. Participants were put through the cold pressor test by immersing their dominant hand above the wrist in water.

Immediately after the CPT is administered, participants completed an HL designed lottery task to measure their risk preferences. Qualtrics^{XM} online survey platform was used to design and recorded the HL lottery task for all the participants.

3.3 Measures

Electrodermal activity (EDA) was recorded to measure participants physiological responses that influenced their risk preferences in the presences of stress. According to Horstick et al. (2018), EDA has become a psychophysiological standard method for the measuring sympathetic activity (p.2). Sympathetic nervous system activates what is commonly known as the fight or flight response. Additionally, EDA reflects external and internal factors that influence psychophysical activations due to the integration of central nervous processes into the sweat gland (Horstick et al., 2018, p.2).

EDA was extracted over two experimental episodes: baseline and treatment. The last two minutes of the baseline were used to reduce the noise in the data being analyzed. The first component of EDA recorded was skin conductivity levels (SCL), which are a tonic component of skin conductivity referred to as the general arousal of a person. The second component was skin conductance responses (SCRs), which are short phasic electrodermal responses that increase within one second after a discrete stimulus. Specifically, non-specific skin conductance responses (NS.SCRs) are used. This is the number of SCRs in a period and are considered a tonic measure because they occur post-stimuli (Posada-Quintero et al., 2016, p.3125).

EDA data was recorded with Biopac MP 150. AcqKnowledge® software was used to make corrections and measure EDA data components. First, EDA waveforms were transformed using a low-pass filter of 10 Hz with a sampling rate of eight (Norris et al., 2007, p.824). Second, a square root function was used to adjust for skew within the recorded sample (Dawson et al., 2000, p.226). Third, to preprocess and delete artifacts in the data in order to identify NS.SCRs, the sampling rate of the recorded waveforms were reduced to 31.25 Hz and smoothed by a median filter (Horstick et al., 2018, p.4).

3.4 Psychological Measures

The main psychological measures being used for this experiment is positive affect and negative affect schedule (PANAS). Watson et al. (1988) designed the PANAS to assess participants' negative and positive affect. Item scores range from 1 "not at all" to 5 "extremely" (p.1063). Positive affect (PA) and Negative affect (NA) subscales were computed by averaging the ten items per subscale (Watson et al, 1988, p.1063). The PANAS was assessed before and after the CPT.

Results

All computations are carried out using the Stata® statistical package (Version 14). The participants were randomly assigned to two groups, the control group and the treatment group. 30 participants were assigned to the control group, and 38 were assigned to the treatment group. Those in the treatment group underwent physiological change due to stress. The dependent variable for this experiment is Risk Aversion (RA) measured within the HL lottery task. ΔSCL and ΔNS.SCRs are the two components of EDA data that are being considered as independent variables.

4.1 Physiological Measures

4.1.1 Skin Conductance Level (SCL)

The change in SCL (ΔSCL) is the difference between the mean SCL during the CPT and the mean SCL during the baseline. For the baseline, the last two minutes are used to create the independent variable to reduce participants basal variability. As a result of baseline adjustments for the participant's (n=27), ΔSCL became negative. This is observed in both groups. A constant k=1 was added to all participants ΔSCL to resolve this issue which is a common practice in data analysis (Dawson et al, 2000, p.226).

The ΔSCL in the control group is normally distributed, however it is not normally distributed in the treatment group (Table 3 & 4). Also, the ΔSCL does not exhibit skewness and kurtosis in the control group; but it does exhibit skewness and kurtosis in treatment group (Table 5 & 6).

The results of an independent t-test showed that participants in the treatment group do have statistically different ΔSCL versus participants in the control group (MC: 0.98 ± 0.15 μS; MT: 1.18

± 0.24 μS; t (66) = -4.02, p = 0.0002) (Table 1). Using a 5% confidence level, we can reject the null hypothesis that the difference in means is statistically different for both groups as well as the null hypothesis that the difference in variances is statistically different for both groups (F(66) = 0.388, p = 0.0101) (Table 2). The ΔSCL one-way ANOVA test showed that there is statistically significant difference between the two groups (F(1,66) = 14.56, p = 0.0003) (Table 3).

4.1.2 Non-Specific Skin Conductivity Responses (NS.SCRs)

The change in non-specific skin conductivity responses (ΔNS.SCRs) is the difference between the number of SCRs per second during CPT and baseline. Specifically, NS.SCRs are the number of SCRs in a period and are considered a tonic measure because they occur post-stimuli (Posada-Quintero et al., 2016, p.3125). As a result of baseline adjustments, the ΔNS.SCRs for 12 participants' turned negative. Furthermore, 11 participants do not have any ΔNS.SCRs given the chosen thresholds. Hein et. al (2011) suggest that the ΔNS.SCRs an amplitude threshold of 0.005 μS since the stimulus includes participants receiving pain (p.3). Trials with delay periods lower than 5 seconds were excluded to avoid contamination from artefacts caused by the pain stimulators (Hein et. al, 2011, p.3). Similar to ΔSCL, a constant k = 1 was added to preserve the dataset (Dawson et al, 2000, p.226).

The ΔNS.SCRs is normally distributed in the control group but not in the treatment group (Table 3 & 4). Also, the ΔNS.SCRs does not exhibit skewness and kurtosis in the control group. In the treatment group, however, skewness and kurtosis are present (Table 5 & 6).

The results of an independent t-test showed that participants in the treatment group do have statistically different ΔNS.SCRs versus participants in the control group (MC: 0.99 ± 0.04 μS; MT: 1.05 ± 0.07 μS; t (66) = -4.06, p = 0.0001) (Table 1). Using a 5% confidence level, we can reject the null hypothesis that the difference in means is statistically different for both groups as well as the null hypothesis that the difference in variances is statistically different for both groups ($F(66) = 0.3169$, $p = 0.0020$) (Table 2). The ΔNS.SCRs one-way ANOVA test showed a statistically significant difference between the two groups ($F(1,66) = 14.58$, $p = 0.0003$) (Table 3).

4.2.1 PANAS—Positive Affect (PA)

The ΔPA in the control group is not normally distributed and it exhibits skewness and kurtosis (Table 3 & 5). In the treatment group, however, the ΔPA is normally distributed and exhibits skewness and kurtosis (Table 4 & 6). The results of an independent t-test showed that the participants in the treatment group do not have statistically different ΔPA versus participants in the control group (MC: -0.8 ± .96 ΔPA; MT: -0.13. ± 0.89 ΔPA; $t(66) = -0.5115$, $p = 0.6108$). Using a 5% confidence level, we cannot reject the null hypothesis that the difference in means is statistically different for both groups as well as the null hypothesis that the difference in variances is statistically different for both groups ($F(66) = 0.9179$, $p = 0.8198$). The ΔPA one-way ANOVA test showed that there is no statistically significant difference between the two groups ($F(1,66) = .89$, $p = 0.5770$).

4.2.2 PANAS—Negative Affect (NA)

The ΔNA in the control group is not normally distributed and exhibits skewness and kurtosis (Table 3 & 5). For the treatment group however, the ΔNA is normally distributed and does not exhibit skewness and kurtosis (Table 4 & 6). The results of an independent t-test showed that participants in the treatment group do not have statistically different ΔNA versus participants in the control group (MC:2.37 ± .63 ΔNA; MT: 1.53 ± .67 ΔNA; $t(66) = 0.9083$, $p = 0.3670$). Using a 5% confidence level, we cannot reject the null hypothesis that the difference in means is statistically different for both groups as well as the null hypothesis that the difference in variances is statistically different for both groups ($f(66) = .6962$, $p = 0.3173$). The ΔNA one-way ANOVA test showed no statistically significant difference between groups with or without stress ($F(1,66) = .94$, $p = 0.5404$).

Table 1: T-test comparing control and treatment groups.

	n	Mean	SD	t-cal	Satterthwaite's df	p	Decision
RA							
Control	30	.54	.54	-.08	50	.938	Accept
Treatment	38	.56	1.38				
# Safe Choices							
Control	30	5.2	1.27	.68	63	.497	Accept
Treatment	38	4.92	2.07				
ΔSCL							
Control	30	.98	.16	-4.02	63	.000	Reject
Treatment	38	1.18	.15				
ΔNS.SCRs							
Control	30	.99	.04	-4.06	60	.000	Reject
Treatment	38	1.05	.07				
ΔPA							
Control	30	-.8	5.25	-.51	64	.610	Accept
Treatment	38	-.13	5.48				
ΔNA							
Control	30	2.37	3.47	.909	66	.367	Accept
Treatment	38	1.53	4.16				

Table 2: Variance Results Comparing Control and Treatment Groups.

	n	F	df	p	Decision
RA					
Control	30	.156	29, 37	.000	Reject
Treatment	38				
# Safe Choices					
Control	30	.376	29, 37	.008	Reject
Treatment	38				
ΔSCL					
Control	30	.388	29, 37	.010	Reject
Treatment	38				
ΔNS.SCRs					
Control	30	.317	29, 37	.002	Reject
Treatment	38				
ΔPA					
Control	30	.917	29, 37	.819	Accept
Treatment	38				
ΔNA					
Control	30	.692	29, 37	.3173	Accept
Treatment	38				

Table 3: One-way ANOVA: Total Risk Aversion

	Sum of Squares	df	Mean Square	F	Sig.	Decision
Risk Aversion						
Between Groups	.006	1	.006	0	.944	Accept
Within Groups	78.8	66	1.19			
Total	78.8	67	1.17			
# Safe Choices						
Between Groups	1.63	1	1.63	.52	.471	Accept
Within Groups	206.1	66	3.12			
Total	207.7	67				
ΔSCL						
Between Groups	.65	1	.643	14.56	.0003	Reject
Within Groups	2.94	66	.045			
Total	3.59	67	.054			
ΔSCRs						
Between Groups	.055	1	.055	14.58	.000	Reject
Within Groups	.249	66	.004			
Total	.304	67	.006			
ΔPA						
Between Groups	7.49	1	7.49	.26	.613	Accept
Within Groups	1909.14	66	28.93			
Total	1916.63	67	28.60			
	Sum of Squares	Df	Mean Square	F	Sig.	Decision
Risk Aversion						
Between Groups	.006	1	.006	0	.944	Accept
Within Groups	78.8	66	1.19			
Total	78.8	67	1.17			
ΔSCL						
Between Groups	.65	1	.643	14.56	.0003	Reject
Within Groups	2.94	66	.045			
Total	3.59	67	.054			
ΔSCRs						
Between Groups	.055	1	.055	14.58	.000	Reject
Within Groups	.249	66	.004			
Total	.304	67	.006			
ΔPA						
Between Groups	7.49	1	7.49	.26	.613	Accept
Within Groups	1909.14	66	28.93			
Total	1916.63	67	28.60			

4.3 Control Variables

Age, Sex, Education levels, and average Income levels are the four control variables being considered. Participant's ages ranged between 18–58, with an average (M) = 24.75 and standard deviation (SD) = 9.87. In the control group, participant's ages ranged between 18–48, M = 22.46 and SD = 6.84. In the treatment group, participant's ages ranged between 18–58, M = 26.55, and SD = 11.5 (Table 7). For the aggregated sample pool, 46% of participants were males and 54% were females. In the control group, 30% of participants were males and 70% were females. In the treatment group, 58% of participants were males and 42% were females (Table 7).

Participants were asked if they consider themselves below average income, average income, and above average income for their age group. 56% of participants categorized themselves as below average income, 40% identified themselves as average income, and 4% identified themselves as above average income. In the control group, 50% of participants categorized themselves as below average income, 43% as average, and 7% as above average. In the treatment group, 61% identified themselves as below average income, 37% as average, and 3% above average (Table 7).

Participants were also asked their highest level of education. For the aggregated sample pool, 16% of participants' received only a high school diploma, 50% of them had some college but no degree, 15% had a bachelor's degree, 13% had a master's degree, and 6% had a doctoral degree. In the control group, 30% of participants' received only a high school diploma, 47% had some college but no degree, 13% had a bachelor's degree, and 10% had a master's degree. In the treatment group, 5% of participants' received only a high school diploma, 53% had some college but no degree, 16% had a bachelor's degree, 16% had a master's degree, and 11% had a doctoral degree (Table 7).

4.4 Distribution Ladder

Tukey's ladder of powers test was used to identify significant functional transformations for the two physiological variables in the treatment group since they are not normally distributed. The Tukey test searches a subset of the ladder of powers for a transformation that converts non-normal variables into normally distributed ones (Tukey, 1949, p.99). Ladder transformation for ΔSCL shows that out of nine possible functional transformations, the inverse (1/ΔSCL) and one over square [$1/(\Delta SCL)^2$] will normalize the data. The inverse of ΔSCL will be used to further the analysis since it has a clearer interpretation as a measure of physiological arousal. For ΔNS.SCRs there is no known function that will normalize the distribution of the variable; however, inverse function will be used to be consistent within the data analysis.

4.5 Correlations Matrixes

Pearson's correlation coefficient matrix was used after non-normally distributed variables were normalized (Horstick et al., 2018, p.4). A correlation matrix is constructed between the dependent variable RA and the independent variables 1/ΔSCL, 1/ΔSCRs, ΔPA, and ΔNA. A second correlation coefficient matrix is between RA and four control variables was also considered: Age, Sex, Education, and average Income levels. Additionally, CPT, a dummy variable for the treatment effect, has been include in the matrix.

RA and Education are statistically significant and negatively correlated (0.26, $p = 0.0354$) (Table 9). Furthermore, CPT has a statically significant negative correlations with 1/ΔSCL (-0.48, $p = 0.0000$) and a negative correlation with 1/ΔNS.SCRs (-0.46, $p = 0.0001$) (Table 8). The two physiological measures 1/ΔSCL and 1/ΔNS.SCRs have a strong positive correlation (0.65, $p = 0.0038$) (Table 8).

There was no significant relationship between the dependent variable RA and the independent variables $1/\Delta SCL$, $1/\Delta SCRs$, ΔPA, and ΔNA (Table 8). In addition, RA does not have a significant relationship with the three control variables: Age, Sex, and average Income levels (Table 9).

4.6 Financial Decisions—Risk Aversion (RA)

RA was calculated for each of the ten HL lottery choices using the utility function $U(x)=x^{\wedge(1-r)}/(1-r)$ for money x, $x > 0$. RA was calculated by solving for r then taking the average of RA for lotteries before and after participants switch. For example, if the switching point was at lottery 7, RA was calculated for lotteries 6 and 7, then the average of the two was used. RA for both groups is normally distributed and does not exhibit skewness and kurtosis (Table 3, 4, 5 & 6).

The results of an independent t-test showed participants in the treatment group do not have statistically different RA versus participants in the control group (MC: 0.54 ± 0.54 r; MT: 0.56 ± 1.37 r; $t(66) = -0.077$, $p = 0.9389$). Using a 5% confidence level, we cannot reject the null hypothesis that the difference in means is statistically different for both groups. However, can reject the null hypothesis that the difference in variances is statistically different for both groups ($F(66) = .1560$, $p = 0.0000$). The RA one-way ANOVA test showed that there was no statistically significant difference between the two groups ($F(1,66) = 0$, $p = 0.944$).

4.7 Number of Safe Choices

As a second dependent variable the number of safe choices was recorded in the HL. The number of safe choices is referring to the number of times participants choosing a less risky option (option A) versus a more risky option (option B). There are ten lottery choices in the HL, hence participants can make a maximum of ten safe choices. The number of safe choices was created as a variable in order to check the robustness of findings using RA as the dependent variable. The results of an independent t-test showed participants in the treatment group do not have statistically different number of safe choices versus participants in the control group (MC: 5.2 ± 0.23 r; MT: $4.92 \pm .33$ r; $t(66) = 0.68$, $p = 0.497$). Using a 5% confidence level, we cannot reject the null hypothesis that the difference in means is statistically different for both groups. However, can reject the null hypothesis that the difference in variances is statistically different for both groups ($F(66) = .388$, $p = 0.010$). The number of safe choices one-way ANOVA test showed that there was no statistically significant difference between the two groups ($F(1,66) = .52$, $p = 0.471$).

Conclusion and Discussion

This experiment is contributing to the gap in literature by adding the results of a laboratory experiment on financial decision making under stress. Financial decision making has been studied vigorously due to its relevance to consumer and producer utility. However, on an individual level, there was a gap in the literature when considering financial decision making and stress. The results of the current body of work within the field of Neuroeconomics provides a solution using physiological measures: ΔSCL and ΔSCRs. In the treatment group, ΔSCL and ΔNS.SCRs are different from the control group, indicating that stress is changing participants physiological state.

Comparing participants means for RA in HL lottery task it is evident there is no statistical difference between both groups. Participants on average have the same risk aversion preferences with or without the occurrences of stress. On average, participants are likely to make the same financial decisions. However, the variances for the two groups are statistically different. More participants in the treatment group are exhibiting risk aversion in sure gains and risk seeking in sure losses than in the control group.

The difference in variances has a similar outcome as described by certainty effect and energy budget rule. Stress is a similar stimulus when it comes to making financial decisions as is certainty. For animals, a similar stimulus occurs when an animal is danger of starvation. A participant's financial decisions under stress are extreme. That is, under stress in HL, more participants are choosing one, two, nine, or ten safe lotteries. Approximately 21% of participants under stress choose the two tail ended lotteries versus around 6% of participants without stress (Table 2).

Participants behaving consistently in their lottery choices was a key assumption when constructing the dependent variable. Within the data, it was observed that participants areswitching from a less risky to a more risky option more than once. As pointed out in the original HL paper, even for those who switched back and forth, there is typically a clear division point between clusters of less risky and more risky choices, with few errors on each side (p.1648). The dependent variable was constructed using this clear division in participants lottery choices. The isolation effect provides a possible explanation in which individuals tend to be inconsistent when presented with same choice in different forms. Specifically, 55% of participants in the treatment group are switching more than once from a less risky (Option A) to a more risky option (Option B) versus only 30% in the control group. That is, a much higher percentage of participants in the treatment group are exhibiting inconsistent financial decisions.

Finally, financial decision making and physiological change as a result of stress are critical to understanding real world financial outcomes. This study illustrates the impact of stress has on financial decision making and concludes that stress plays a vital role in changing participants' risk preferences.

References

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Chicago (Ill.): University of Chicago Press.*

Bhanji, J. P., Kim, E. S., & Delgado, M. R. (2016). Perceived control alters the effect of stress on persistence. *Journal of Experimental Psychology: General, 145*(3), 356-365. doi:10.1037/xge0000137

Cohen, A. & Einav, L. (2007). Estimating Risk Preferences from Deductible Choice. *American Economic Review 97*(3), 745-88. doi:10.1257/aer.97.3.745

Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization, 87*, 43-51. doi:10.1016/j.jebo.2012.12.023

Dawson, E. M., Schell, M. A., & Filion, L. D. (2000). The electrodermal system. *Handbook of Psychophysiology*. doi: 10.1017/CBO9780511546396.007

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43–74. doi:10.1016/0010-0277(92)90050-R

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504-528. doi:10.1016/s0092-6566(03)00046-1

Hines, E. A., & Brown, G. E. (1936). The cold pressor test for measuring the reactibility of the blood pressure: Data concerning 571 normal and hypertensive subjects. *American Heart Journal, 11*(1), 1-9. doi:10.1016/s0002-8703(36)90370-8

Hein, G., Lamm, C., Brodbeck, C., & Singer, T. (2011). Skin Conductance Response to the Pain of Others Predicts Later Costly Helping. *PLoS ONE,* 6(8). doi:10.1371/journal.pone.0022759

Holt, C. A., & Laury, S. (2002). Risk Aversion and Incentive Effects. *SSRN Electronic Journal*. doi:10.2139/ssrn.893797

Horstick, J., Siebers, S., & Backhaus, C. (2018). Examination of the Skin Conductance Level (SCL) as an Index of the Activity of the Sympathetic Nervous System for Application in the Continous Blood Pressure Measurement. *International Journal of Biomedical Engineering and Science, 5*(2), 01-11. doi:10.5121/ijbes.2018.5201

Kahneman, D., & Tversky, A. (1979). Prospect Theory. An Analysis of Decision Making Under Risk. *Econometrica,* doi:10.21236/ada045771

Neumann, J. V., & Morgenstern, O. (1953). *Theory of games and economic behavior*. Princeton: Princeton Univ. Pr.

Norris, C. J., Larsen, J. T., & Cacioppo, J. T. (2007). Neuroticism is associated with larger and more prolonged electrodermal responses to emotionally evocative pictures. *Psychophysiology, 44*(5), 823-826. doi:10.1111/j.1469-8986.2007.00551.x

Tukey, J. W. (1949). Comparing Individual Means in the Analysis of Variance. *Biometrics, 5*(2), 99. doi:10.2307/3001913

Ruixun, Z., Brennan, J. T., & Lo W. A. (2014). The Origin of Risk Aversion. *Proceedings of the National Academy of Sciences 111*, (50), 17777-7782. doi:10.1073/pnas.1406755111

Watson, D., Clark, L. A., & Tellegen, A. (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. doi:10.1037/0022-3514.54.6.1063

Weber, U. E., Shafir, S., & Blais, R. A. (2004). Predicting Risk Sensitivity in Humans and Lower Animals: Risk as Variance or Coefficient of Variation. *Psychological Review 111*(2), 430-45. doi:10.1037/0033-295x.111.2.430

Appendix A

Table 1: This Holt Laury lottery task was presented to the participants. To aid in explaining the lotteries, they have been rearranged with the highest expected payoff differences from choosing the less risky option being on top.

	Option A	Option B	Expected Payoff Differences
1	10% chance of winning \$4.00 and 90% chance of winning \$3.50	10% chance of winning \$7.00 and 90% chance of winning \$1.00	\$1.95
2	20% chance of winning \$4.00 and 80% chance of winning \$3.50	20% chance of winning \$7.00 and 80% chance of winning \$1.00	\$1.40
3	30% chance of winning \$4.00 and 70% chance of winning \$3.50	30% chance of winning \$7.00 and 70% chance of winning \$1.00	\$.85
4	40% chance of winning \$4.00 and 60% chance of winning \$3.50	40% chance of winning \$7.00 and 60% chance of winning \$1.00	\$.30
5	50% chance of winning \$4.00 and 50% chance of winning \$3.50	50% chance of winning \$7.00 and 50% chance of winning \$1.00	-\$.25
6	60% chance of winning \$4.00 and 40% chance of winning \$3.50	60% chance of winning \$7.00 and 40% chance of winning \$1.00	-\$.80
7	70% chance of winning \$4.00 and 30% chance of winning \$3.50	70% chance of winning \$7.00 and 30% chance of winning \$1.00	-\$1.35
8	80% chance of winning \$4.00 and 20% chance of winning \$3.50	80% chance of winning \$7.00 and 20% chance of winning \$1.00	-\$1.90
9	90% chance of winning \$4.00 and 10% chance of winning \$3.50	90% chance of winning \$7.00 and 10% chance of winning \$1.00	-\$2.45
10	100% chance of winning \$4.00 and 0% chance of winning \$3.50	100% chance of winning \$7.00 and 0% chance of winning \$1.00	-\$3

Table 2: Calculated risk aversion within the HL designed lottery task.

Number of Safe Choices	Range of Relative Risk Aversion U(x)=x^(1-r)/(1-r)	Risk Preference Classifications	Proportion of Choices for Treatment Group	Proportion of Choices for Control Group
0-1	r < -1.37	Highly Risk Loving	8%	3%
2	-1.37 < r < -.74	Very Risk Loving	0%	0%
3	-.74 < r < -.24	Risk Loving	11%	13%
4	-.24 < r < .19	Risk Neutral	16%	7%
5	.19 < r < .61	Slightly Risk Averse	24%	23%
6	.61 < r < 1.04	Risk Averse	16%	43%
7	1.04 < r < 1.54	Very Risk Averse	9%	7%
8	1.54 < r < 2.26	Highly Risk Averse	5%	3%
9-10	2.26 < r	Stay in Bed	8%	0%

Table 3: Shapiro-Wilk test was used to determine if the variables are normally distributed for the control group, n=30.

	W	V	Z	Prob > z
Risk Aversion	.98034	.625	-.972	.8344
ΔSCL	.97372	.835	-.372	.6449
ΔSCRs	.97334	.847	-.342	.6339
ΔPANAS pos	.90097	3.148	2.371	.0088
ΔPANAS neg	.88130	3.773	2.746	.0030

Table 4: Shapiro-Wilk test was used to determine if the variables are normally distributed for the treatment group, n=38.

	W	V	Z	Prob > z
Risk Aversion	.94628	2.041	1.497	.0672
ΔSCL	.82620	6.718	3.996	.0000
ΔSCRs	.75093	9.463	4.715	.0000
ΔPANAS pos	.94888	1.943	1.393	.0817
ΔPANAS neg	.98621	.524	-1.355	.9123

Table 5: Jarqu-Berra is a goodness of fit test was used to determine if variables are exhibiting Skewness and Kurtosis for the control group, n =30.

	Pr (Skewness)	Pr (Kurtosis	Adj chi2	Prob > chi2
Risk Aversion	.7613	.4345	.74	.6913
ΔSCL	.1192	.2900	3.83	.1476
ΔSCRs	.6098	.9545	.26	.8765
ΔPANAS pos	.0082	.0022	12.73	.0017
ΔPANAS neg	.0053	.1636	8.29	.0158

Table 6: Jarqu-Berra is a goodness of fit test was used to determine if variables are exhibiting Skewness and Kurtosis for the treatment group, n =38.

	Pr (Skewness)	Pr (Kurtosis	Adj chi2	Prob > chi2
Risk Aversion	.8758	.1319	2.46	.2922
ΔSCL	.0000	.0007	20.66	.0000
ΔSCRs	.0000	.0001	27.14	.0000
ΔPANAS pos	.0224	.0172	9.16	.0103
ΔPANAS neg	.5103	.1840	2.36	.3079

Table 7: Participant demographics anomalously collected and stored.

	Control (n=30)	Treatment (n=38)	Total (n=68)
Male	30%	58%	46%
Female	70%	42%	54%
Range Age	18-48	18-58	18-58
Average Age	22.46	22.46	24.75
Standard Deviation Age	6.84	11.5	9.87
Below Average Income	50%	61%	56%
Average Income	43%	37%	40%
Above Average Income	7%	3%	4%
High School Graduated	30%	5%	16%
Some College No Degree	47%	53%	50%
Bachelor’s Degree	13%	16%	15%
Master’s Degree	10%	16%	13%
Doctoral Degree	0%	11%	6%

Table 8: Correlation matrix was constructed between the dependent variable and the independent variables that include the two physiological measures ΔSCL and ΔSCRs; note the inverse function was used to normalize the independent variables. Also, two psychological measures ΔPA and ΔNA were used, and a dummy variable as a proxy for treatment effect called CPT. There are 68 participants.

	Risk Aversion	Inv(ΔSCRs)	Inv(ΔSCL)	ΔPA	ΔNA	CPT
Risk Aversion	1.0000					
Inv(ΔSCRs)	-0.1026 0.4050	1.0000				
Inv(ΔSCL)	-0.1070 0.3852	0.6483* 0.0000	1.0000			
ΔPA	-0.1165 0.3443	-0.0715 0.5621	-0.2042 0.0948	1.0000		
ΔNA	0.0134 0.9135	0.0216 0.8614	0.0331 0.1884	-0.0700 0.5703	1.0000	
CPT	0.0087 0.9440	-0.4569* 0.0001	-0.4765* 0.0000	0.0625 0.6125	-0.1088 0.3772	1.0000

Table 9: Correlation matrix was constructed between the dependent variable and four control variables: Age, Sex, Education levels, and average Income levels Also, a dummy variable as a proxy for treatment effect called CPT was included. There are 68 participants.

	Risk Aversion	Age	Sex	Education	Income	CPT
Risk Aversion	1.0000					
Age	-0.1565 0.2024	1.0000				
Sex	0.1751 0.1532	0.0050 0.6561	1.0000			
Education	-0.2555* 0.0354	0.5722* 0.0000	0.0602 0.6257	1.0000		
Income	0.0505 0.6824	-0.1645 0.1800	0.0022 0.9855	-0.2340* 0.0548	1.0000	
CPT	0.0087 0.9940	0.2070 0.0903	-0.2781* 0.0217	0.3208* 0.0077	-0.1244 0.3123	1.0000

Chapter 2

Inconsistent Financial Decision Making Under Stress

Introduction

This research paper presents new evidence which proves that participants under stress make inconsistent financial decisions. Participants made inconsistent financial decisions in a laboratory experiment that included 30 participants in the control group, and 38 participants in the treatment group. In the treatment group, participants' financial decisions were inconsistent at a much higher overall percentage, compared to the control group without stress.

The Holt Laury's (HL) lottery task is used to assess participants' risk aversion. The Cold Pressor Test (CPT) is used to induce a safe level of stress in the treatment group. In the HL lottery task, participants completed ten lottery choices between less and more risky options, allowing risk aversion to be calculated. For participants' risk aversion to be calculated, a participant who first chooses the less risky option should switch only once from a less risky to a more risky option. That is, once a participant chooses a more risky lottery, she is expected never to switch again to a less risky option. Any participant that would depart from this behavioral pattern would be exhibiting inconsistent financial decision making.

One possible solution when working with participants' inconsistent data is to look at the clear division in the data, as pointed out in the original HL paper. Specifically, for participants who are inconsistent, there is typically a clear division point between clusters of less and more risky options, with few errors on each side (Holt and Laury, 2002, p.1648). However, in order to gain further insight a dependent variable representing inconsistent financial behavior is created and analyzed. Participants' physiological and psychological measures are used in order to explain their inconsistent financial decisions.

Literature Review

2.1 Holt and Laury

Holt and Laury (HL) in 2002 introduced a menu of choices tasks that can be used to estimate the degree of risk aversion as well as specific functional form. Participants in an HL designed experiment choose between two options ten times as the expected payoffs change (Holt and Laury, 2002, p.1644). Participants should switch options once the magnitude of expected payoffs changes from positive to negative (Holt and Laury, 2002, p.1645). Depending on when participants switch options, their risk preference is revealed (Holt and Laury, 2002, p.1646).

For the current experiment, the payoffs for the less risky option is \$4.00 or \$3.50 (option A), which is much less than the potential payoffs of \$7.00 or \$1.00 in the more risky option (option B) (Table 1). The two options are randomly presented to participants ten times. Each time the options are presented, the probability of earning a higher payoff increases and the probability of earning a lower payoff decreases. For example, in the first choice, the probability of the high payoff (\$4.00 or \$7.00) for each option is 1/10 versus for the lower payoff (\$3.50 or \$1.00) the probability is 9/10; hence, only an extreme risk seeker would choose a more risky option (Option B) (Table 1). On the other hand, even the most risk-averse person should switch over by decision 10 in the bottom row, since the more risky option (option B) yields a sure payoff of \$7.00 (Table 1).

The switching point is determined where expected payoff differences changes from positive to negative. Specifically, the expected payoff incentive to choose less risky Option A ranges from \$1.95 to -\$3.00 (Table 1). The payoffs for the lottery choices in the experiment were selected so that the switching point would provide an interval estimate of a participant's coefficient of relative risk aversion (Table 2). Furthermore, the payoff numbers for the lotteries so that the risk-neutral choice pattern of four safe choices followed by six risky choices was optimal for constant relative risk aversion in the interval (-.24,.19) (Table 2).

In literature, constant relative risk aversion is generally assumed due to functional form being logarithmic linear and computationally convenient (Holt and Laury, 2002, p.1646). Therefore, calculated risk aversion can be interpreted as risk preference for $r < 0$, risk neutrality for $r = 0$, and risk aversion for $r > 0$ (Holt and Laury, 2002, p.1646). However, constant relative risk aversion is not a necessary assumption for HL designed experiments.

2.2 Cold Pressor Test

Hines and Brown (1936) originally developed the Cold Pressor Test (CPT) as a procedure which is carried out by immersing an extremity in ice water (p.1). That is, one hand was immersed above the wrist in ice water ranging from 4^{O} to 5^{O} C for about 20 to 30 seconds (Hines & Brown, 1936, p.2). In recent literature, experimentalists have used two components of electrodermal activity (EDA) to record and analyze participants' responses to the CPT: skin conductance level (SCL) and non-specific skin conductance responses (NS.SCRs).

For example, the CPT was used in an experiment in which SCLs were utilized to develop a relative intra-individual comparison (Horstick et al., 2018, p.2). Specifically, authors demonstrated that the SCL can be used as an index of sympathetic activation as result of the effects of physical stress situations (Horstick et al., 2018, p.8). Also, Posada-Quintero et al. (2016) use EDA differences between baseline conditions and three treatments to elicit sympathetic activation: postural stimulation, CPT, and the Stroop test (p.3125). The authors found significant differences in NS.SCRs and SCL when comparing baseline to CPT (Posada-Quintero et al., 2016, p.3125).

2.3 Inconsistent Financial Decisions

Consistency is the fundamental axiom underlying rational behavior in the neoclassical construct of utility maximization (Afriat, 1976, p.76). The study of consistency was formalized with the Generalized Axiom of Revealed Preference (GARP), which is equivalent to Afriat's cyclical consistency condition (Varian, 1982, p.947). For GARP to be satisfied, it is a necessary and sufficient condition for data to be consistent with utility maximization (Varian, 1982, p.948).

Regardless of the popoularity of GARP in Neoclassical Economics, numerous studies have provided examples of it being violated. Within the field of Behavioral Psychology, evidence of inconsistency was provided by how choices are framed (Kamen and Taversky, 1981, p.453). Inconsistent responses to problems arise from the conjunction of a framing effect with contradictory attitudes toward risks involving gains and losses (Kamen and Taversky, 1981, p.453). Also, the isolation effect describes individuals with the tendency to discard components that are shared by all prospects under consideration (Kahneman & Tversky, 1979, p.271). The isolation effect leads to inconsistent preferences when the same choice is presented in different forms (Kahneman & Tversky, 1979, p.271).

Recent laboratory experiments suggest that a degree of choice inconsistency might be present in human decision making. Specifically, HL method allows participants to switch freely between options, which may lead to participants making inconsistent choices by switching more than once (Charness et al., 2013, p.47). For example, 25% of participants switched their lottery choices more than once in a study that compares the HL's risk attitude elicitation with a risk attitude classification associated with insurance behavior (Corcos et al., 2018, p.13).

Methods

3.1 Sample

Potential participants were recruited utilizing an approved email distribution list from the Center for Neuroeconomics Studies (CNS) website. In total, 68 individuals were recruited to participate in the experiment titled Risk Experiment Under Stress. The laboratory experiment was approved by the Institutional Review Board (#2914). Participants were notified before arriving that the opportunity involved completing surveys, some of which would be done after immersing their dominant hand in cold or warm water. They were also informed that immersing their hand would increase their heart rate and cause them varying levels of pain.

3.2 Procedure

Participants were asked to arrive at the Center for Neuroeconomics Studies (CNS) lab located in Claremont Graduate University, Claremont, CA. Upon arrival, participants were given a consent form with a detailed explanation of the experiment. Within the consent form, a clear description of the earning was disclosed: "You will earn a maximum of $7 on the lottery task that you will complete after you immerse your hand in ice or warm water. You will earn $.50 for every 15 secs of you holding your hand under ice or warm water, maximum amount of time allowed is 2 minutes. Possible earnings total $11" (Consent Form).

Once a written consent was provided, participants received an identity masking code and were randomly assigned to control or treatment groups. Participants were informed their participation would take approximately 25-30 minutes. Immediately, a lab administrator screened for conflicts with increased heart rate due to the cold pressor test. Participants were excluded if they had a history of cardiovascular disorder, fainting, seizures, Reynald's phenomenon, frostbite, or an open cut/fracture on their dominant hand. Also, participants under the age of 18 were excluded.

Once written consent was provided, participants received an identity masking code and were randomly assigned to control or treatment groups. Participants were informed their participation would take approximately 25-30 min. Immediately, a lab administrator screened for conflicts with increased heart rate due to the cold pressor test. Participants were excluded if they had a history of cardiovascular disorder, fainting, seizures, Raynaud's phenomenon, frostbite, and open cut/fracture on their dominant hand. Also, participants under the age of 18 were excluded.

Verbal instruction was given before beginning the experiment to assure participants understood their earnings and the CPT procedure. Participants were informed that they may remove their hand from the water when they wished. Furthermore, they should keep in mind the longer they kept their hand in the water the higher their pay will be. Participants were informed that a maximum of two minutes was allowed and the experimenter would stop the CPT only when they have reached the maximum of two minutes. Participants were not informed of how long they had their hand in water, only when they reached the maximum amount of time allowed.

The only difference between the control and treatment groups was the water temperature. The temperature of the water was recorded a few minutes prior to participants scheduled time slot. However, the temperature was not disclosed to the participants. Hence, participants had no prior knowledge if they were in the control versus treatment groups before the beginning of the experiment. Warm water temperature for the control group ranged between 20.3^{O} and 20.5^{O} C; which was approximately room temperature. The cold water ranged between -2^{O} and $-.6^{O}$ C.

Thereafter, the experimenter prepared and attached electrodermal activity (EDA) sensors onto participants non-dominate hand. Once EDA sensors were placed, participants were asked to sit and rest for three minutes while baseline levels were recorded. Participants were put through the cold pressor test by immersing their dominant hand above the wrist in water.

Immediately after the CPT is administered, participants completed an HL designed lottery task to measure their risk preferences. Qualtrics^{XM} online survey platform was used to design and recorded the HL lottery task for all the participants.

3.3 Measures

Electrodermal activity (EDA) was recorded to measure participants' physiological responses influenced their risk preferences in the presences of stress. According to Horstick et al. (2018), EDA has become a psychophysiological standard method for measuring sympathetic activity (p.2). The sympathetic nervous system activates what is commonly known as the fight or flight response. Additionally, EDA reflects external and internal factors that influence psychophysical activations due to the integration of central nervous processes into the vegetative sweat gland (Horstick et al., 2018, p.2).

EDA was extracted over two experimental episodes: baseline and treatment. The last two minutes of the baseline were used to reduce the noise in the data being analyzed. The first component of EDA recorded was skin conductivity levels (SCL), which are a tonic component of skin conductivity referred to as the general arousal of a person. The second component was skin conductance responses (SCRs), which are short phasic electrodermal responses that increase within one second after a discrete stimulus. Specifically, non-specific skin conductance responses (NS.SCRs) are used. This is the number of SCRs in a period and are considered a tonic measure because they occur post-stimuli (Posada-Quintero et al., 2016, p.3125).

EDA data was recorded with Biopac MP 150. AcqKnowledge® software was used to make corrections and measure EDA data components. First, EDA waveforms were transformed using a low-pass filter of 10 Hz with a sampling rate of eight (Norris et al., 2007, p.824). Second, a square root function was used to adjust for skew within the recorded sample (Dawson et al., 2000, p.226). Third, to preprocess and delete artifacts in the data in order to identify NS.SCRs, the sampling rate of the recorded waveforms were reduced to 31.25 Hz and smoothed by a median filter (Horstick et al., 2018, p.4).

3.4 Variables

The dependent variable is the number of switches from a less risky to a more risky option in the HL lottery task. To gain insight into inconsistent financial behavior, the dependent variable Inconsistent was created. Inconsistent is an ordinal dependent variable that represents whether participants are switching once, twice, or three times in the HL. The two main independent variables being considered are the change in SCL (ΔSCL) and the change in NS.SCRs (ΔNS.SCRs). Specifically, the results of a laboratory experiment showed participants do have statistically different ΔSCL and ΔNS.SCRs as result of stress (Bejanyan, 2020, p.11).

The hypothesis for this study is that as result of stress being prevalent in the treatment group, a larger percentage of participants will make inconsistent financial decisions. To establish causation, the treatment was designed to increase participants' level of stress during financial decision making. Outcomes are compared between the treatment and control groups to accumulate an understanding of why participants are switching more than once from a less risky to a more risky option in the HL.

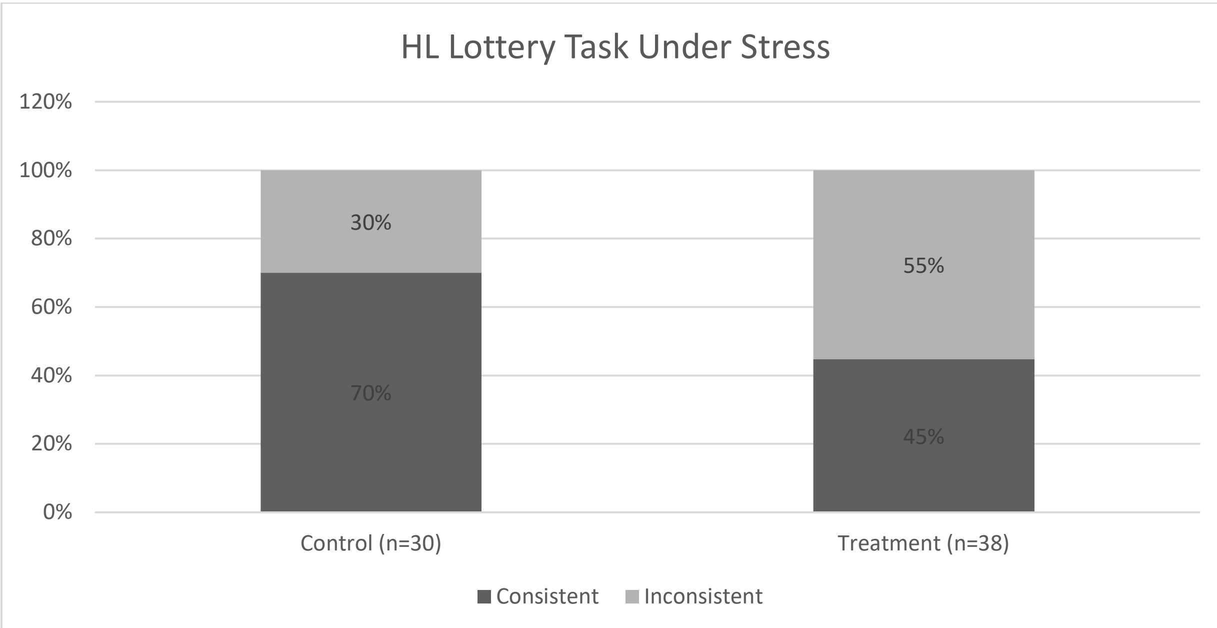
Results

All computations are carried out using the Stata® statistical package (Version 14). Participants without stress were randomly assigned to the control group, n=30. Participants with stress were randomly assigned to the treatment group, n=38. The data for both groups has been combined into one utilizing a dummy variable called CPT which is an indicator for the treatment condition.

The dependent variable Inconsistent is the number of switches from a less risky to a more risky option in the HL lottery task. In the HL, participants are expected to behave rational in their risk preferences. That is, when participants are choosing between two options ten times, they are considered behaving rational if they only switch once from a less risky and a more risky options. Figure 1 shows 55% of participants in the treatment group are switching more than once from a less to a more risky options versus only 30% in the control group. A one-sample t-test was run to determine whether Inconsistent for 68 participants was different than normal, defined as an Inconsistent of one (1.57 ± 0.08, 95% CI, 1.39 to 1.74; $t(67) = 6.578$, $p = 0.000$).

Two components of EDA data that are being considered as independent variables are ΔSCL and ΔNS.SCRs (defined above). Positive Affect (PA) and Negative Affect (NA) two psychological measure that comprise PANAS are being explored. Also, four control variables are being utilized: Age, Sex, Education levels, and Income.

Figure 1: represents the percentage of participants switching more than once from a less risky to a more risky option in the HL lottery task.



4.1 Physiological Measures

4.1.1 Skin Conductance Level (SCL)

The change in SCL (ΔSCL) is the difference between the mean SCL during the CPT and the mean SCL during the baseline. For the baseline, the last two minutes are used to create the independent variable to reduce participants basal variability. As a result of baseline adjustments for the participant's (n=27), ΔSCL became negative. This is observed in both groups. A constant k=1 was added to all participants ΔSCL to resolve this issue which is a common practice in data analysis (Dawson et al, 2000, p.226).

A one-sample t-test was run to determine whether the ΔSCL for 68 participants was different than normal, defined as a ΔSCL of zero (1.10 ± 0.23 µS, 95% CI, 1.04 to 1.15 µS; $t(67) = 39.034$, $p = 0.000$). The Shapiro-Wilk test determined that ΔSCL is not normally distributed, and it exhibits skewness and kurtosis.

Tukey's ladder of powers test was used to identify significant functional transformations the physiological variable since it is not normally distributed. The Tukey test searches a subset of the ladder of powers for a transformation that converts non-normal variables into normally distributed ones (Tukey, 1949, p.99). Ladder transformation for ΔSCL shows that out of nine possible functional transformations, the inverse (1/ΔSCL) and one over square [$1/(\Delta SCL)^2$] will normalize the data. The inverse of ΔSCL will be used to further the analysis since it has a clearer interpretation as a measure of physiological arousal.

4.1.2 Non-Specific Skin Conductivity Responses (NS.SCRs)

The change in non-specific skin conductivity responses (ΔNS.SCRs) is the difference between the number of SCRs per second during CPT and baseline. Specifically, NS.SCRs are the number of SCRs in a period and are considered a tonic measure because they occur post-stimuli (Posada-Quintero et al., 2016, p.3125). As a result of baseline adjustments, the ΔNS.SCRs for 12 participants' turned negative. Furthermore, 11 participants do not have any ΔNS.SCRs given the chosen thresholds. Hein et. al (2011) suggest that the ΔNS.SCRs an amplitude threshold of .005 µS since the stimulus includes participants receiving pain (p.3). Trials with delay periods lower than 5 seconds were excluded to avoid contamination from artefacts caused by the pain stimulators (Hein et. al, 2011, p.3). Similar to ΔSCL, a constant k=1 was added to preserve the dataset (Dawson et al, 2000, p.226).

A one-sample t-test was run to determine whether the ΔNS.SCRs for 68 participants were different than normal, defined as a ΔNS.SCRs of zero (1.03 ± 0.01 µS, 95% CI, 1.01 to 1.05 µS; $t(67)=125.96$, $p=.000$). The Shapiro-Wilk test shows ΔNS.SCRs does not normally distributed, and it exhibits skewness or kurtosis. A ladder transformation for ΔNS.SCRs shows that out of nine possible functional transformations, none will normalize the data. For ΔNS.SCRs there is no known function that will normalize the distribution of the variable; however, an inverse function will be used to be consistent within the data analysis.

4.2 Psychological Measures

The main psychological measures being used for this experiment is positive affect and negative affect schedule (PANAS). Watson et al. (1988) designed the PANAS to assess participants' negative and positive affect. Item scores range from 1 "not at all" to 5 "extremely" (p.1063). Positive affect (PA) and Negative affect (NA) subscales were computed by averaging the ten items per subscale (Watson et al, 1988, p.1063). The PANAS was assessed before and after the CPT. Both PA and NA were baselines adjusted for all participants. The ΔPA for 68 participants has a mean of .43, range $\pm$ .65, with 95% CI, -.87 to 1.72. The ΔNA for 68 participants has a mean of 1.90, range $\pm$ .47, with 95% CI, .96 to 2.83.

4.3 Control Variables

Age, Sex, Education levels, and average Income levels are the four control variables being analyzed. For the aggregated sample pool, 46% of participants were males and 54% were females (Table 3). Participants' ages ranged between 18–58, with an average (M) = 24.75 and standard deviation (SD) = 9.87(Table 3).

Participants were asked if they consider themselves below average income, average income, and above average income for their age group. 56% of participants categorized themselves as below average income, 40% identified themselves as average income, and 4%

identified themselves as above average income (Table 3). Participants were also asked their highest level of education. 16% of participants' received only a high school diploma, 50% of them had some college but no degree, 15% had a bachelor's degree, 13% had a master's degree, and 6% had a doctoral degree (Table 3).

4.4.3 Correlations Matrixes

Pearson's correlation coefficient matrix was used after non-normally distributed variables were normalized (Horstick et al., 2018, p.4). A correlation matrix is constructed between the dependent variable RA and the independent variables 1/ΔSCL, 1/ΔSCRs, ΔPA, and ΔNA. A second correlation coefficient matrix is between RA and four control variables was also considered: Age, Sex, Education, and average Income levels. Additionally, CPT, a dummy variable for the treatment effect, has been include in the matrix.

Inconsistent and CPT are statistically significantly and positively correlated ($r = .2991$, $p = 0.0132$) (Table 4). Furthermore, CPT has statically significant negative correlations with 1/ΔSCL ($r = -.4765$ $p = 0.0000$), and negative correlation1/ΔNS.SCRs ($r = -.4569$, $p = 0.0001$) (Table 4). The two physiological measures 1/ΔSCL and 1/ΔNS.SCRs have a strong positive correlation ($r = .6483$, $p = 0.0038$) (Table 4).

There was no significant correlation between the dependent variable Inconsistent: 1/ΔSCL ($r = .15$, $p = 0.233$); or 1/ΔNS.SCRs ($r = .16$, $p = 0.199$); or ΔPA ($r = -.05$, $p = 0.698$); or ΔNA ($r = -0.18$, $p = 0.136$) (Table 4). In addition, Inconsistent does not have a significant correlation with the four control variables: age ($r = .05$, $p = 0.661$); or sex ($r = -.17$, $p = 0.154$); or education ($r = .22$, $p = 0.078$); or income ($r = .07$, $p = 0.551$) (Table 5).

4.4 Ordered Logistic Regressions

Ordered logistic regressions are estimated to examine the effect stress has on inconsistent financial decision making in the HL. First, a logit model represents Inconsistent as the dependent variable and the independent variables as the Inv(ΔSCL), Inv(ΔNS.SCRs), and CPT. Note that only Inv(ΔSCL) was used in the first logit model, not Inv(ΔNS.SCRs), since these two physiological measures are highly positively correlated with each other ($r = .65$, $p = 0.0038$) (Table 4). The use of both independent variables would result in multicollinearity being prevalent within the logit model, violating one of model's critical assumptions. The results from the logit model show Inconsistent is statistically significantly predicted by Inv(ΔSCL) ($p = .007$) and CPT ($p = .001$) (Table 6).

The second logit model represents Inconsistent as the dependent variable and the independent variables as the Inv(ΔSCL), CPT, ΔPA, and ΔNA. The results from the logit model show Inconsistent financial behavior is statistically significantly predicted by Inv(ΔSCL) ($p = .009$) and CPT ($p = .001$) (Table 6). However, the results from the logit model show Inconsistent is not statistically significantly predicted by ΔPA ($p = 0.646$) and ΔNA ($p = .098$) (Table 6).

The third logit model represents Inconsistent as the dependent variable and the independent variables are Inv(ΔSCL), CPT, ΔPA, ΔNA, Age, and Sex. Note that participants' income and education are negatively correlated ($r = -.23$, $p = 0.055$) (Table 5). Also, participants' age and education are highly positively correlated ($r = .57$, $p = 0.000$) (Table 6), hence only age will be used in the third equation. The results from the logit model show Inconsistent is statistically significantly predicted by Inv(ΔSCL) ($p = 0.008$) and CPT ($p = .002$) (Table 6). However, the results from the logit model show Inconsistent is not statistically significantly predicted by ΔPA ($p = .685$), ΔNA ($p = .192$), Age ($p = .665$), and Sex ($p =. 529$) (Table 6). Additionally, average marginal effects after estimation of ordered logit models were utilized to gain a richer interpretation of the coefficients (Table 7).

Ordered logistic regressions are estimated to examine the effect stress has on inconsistent financial decision making in the HL for the second time using Inv(ΔNS.SCRs). The same three models are analyzed as before however Inv(ΔNS.SCRs) are used only since there is a strong correlation between the two physiological measures (Table 6). The results from the three logit models show Inconsistent financial behavior is statistically significantly predicted by Inv(ΔNS.SCRs) ($p = .007$) and CPT ($p = .001$) (Table 8). However, the results from the logit model show Inconsistent is not statistically significantly predicted by ΔPA ($p = .406$), ΔNA ($p = .178$), Age ($p = .844$), and Sex ($p = .669$) (Table 8). Additionally, average marginal effects after estimation of ordered logit models were utilized to gain a richer interpretation of the coefficients (Table 9).

Conclusion and Discussion

The processes of making financial decisions innately includes individuals making inconsistent decisions. The results from this experiment show in the control group, 30% of participants switched more than once from a less risky to a more risky option in the HL. The inconsistent behavior recorded in the control group is consistent with previous literature (Corcos et al., 2018, p.13).

More importantly, as a result of increasing participants' stress levels, a much higher percentage of participants switched more than once from a less to a more risky option in the treatment group. Specifically, the results show 55% of participants under stress switched more than once. Due to the increased stress the inconsistent behavior normally observed in the HL lottery task has been drastically amplified.

Physiological and psychological data was recorded to substantiate the impact stress has on participants' financial decisions. In particular, the result from the ordered logistic models shows that the number of switches from a less risky to a more risky option is significantly predicted by the Inv(ΔSCL) and CPT. The findings are supported by the work of Posada-Quintero et al., where significant differences in SCL where found when comparing CPT to baseline (Posada-Quintero et al., 2016, p.3125). However, the mediating and control variables do not significantly predicting inconsistent financial behavior: ΔPA, ΔNA, Age, and Sex.

One possibility for solving the issues with participants' inconsistent financial behavior in the HL is to present all ten lottery options at once. In doing so, participants would be forced to be consistent in their choices. The obvious problem with this method is that participants' decisions will not reflect their true preferences. A second solution is provided by Holt and Laury, who recommend looking at the clear division in choices when there are a few errors on each side of the switching point. When considering both possible solutions a key assumption is that participants are making consistent financial decisions, which is a contradicting assumption given the evidence from this experiment.

Finally, this paper adds to the understanding of this obvious disparity in the literature by presenting outcomes from a stress induced experiment. It highlights the need to not ignore inconsistent financial decisions, and in doing so, to not assume individuals are merely behaving irrationally. Specifically, with the use of physiological measures, this paper provides evidence that participants switch more than once in the HL lottery task because of stress.

References

Afriat, S. N. (1967). The construction of utility functions from expenditure data. *New Haven: Cowles Foundation for Research in Economics at Yale University*.

Bejanyan, V. N. (2020) Financial Decision Making Under Stress. *Working paper*.

Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization, 87*, 43-51. doi:10.1016/j.jebo.2012.12.023

Corcos et al. (2018). Measuring individual risk-attitudes: an experimental comparison between Holt & Laury measure and an insurance-choices-based procedure. *Working paper*.

Hines, E. A., & Brown, G. E. (1936). The cold pressor test for measuring the reactibility of the blood pressure: Data concerning 571 normal and hypertensive subjects. *American Heart Journal, 11*(1), 1-9. doi:10.1016/s0002-8703(36)90370-8

Holt, C. A., & Laury, S. (2002). Risk Aversion and Incentive Effects. *SSRN Electronic Journal*. doi:10.2139/ssrn.893797

Horstick, J., Siebers, S., & Backhaus, C. (2018). Examination of the Skin Conductance Level (SCL) as an Index of the Activity of the Sympathetic Nervous System for Application in the Continous Blood Pressure Measurement. *International Journal of Biomedical Engineering and Science, 5*(2), 01-11. doi:10.5121/ijbes.2018.5201

Kahneman, D., & Tversky, A. (1979). Prospect Theory. An Analysis of Decision Making Under Risk. *Econometrica,* doi:10.21236/ada045771

Kahneman, D., & Tversky, A. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458. doi:10.1126/science.7455683

McCullagh, P. (1980). Regression Models for Ordinal Data. *Royal Statistical Society Publications.*

Varian, H. R. (1982). The Nonparametric Approach to Demand Analysis. *Econometrica, 50*(4), 945. doi:10.2307/1912771

Appendix B

Table 1: The Holt Laury's lottery task presented to the participants. To aid in explaining the lotteries, they have been rearranged with the highest expected payoff differences from choosing the less risky option being on top.

	Option A	Option B	Expected Payoff Differences
1	10% chance of winning \$4.00 and 90% chance of winning \$3.50	10% chance of winning \$7.00 and 90% chance of winning \$1.00	\$1.95
2	20% chance of winning \$4.00 and 80% chance of winning \$3.50	20% chance of winning \$7.00 and 80% chance of winning \$1.00	\$1.40
3	30% chance of winning \$4.00 and 70% chance of winning \$3.50	30% chance of winning \$7.00 and 70% chance of winning \$1.00	\$.85
4	40% chance of winning \$4.00 and 60% chance of winning \$3.50	40% chance of winning \$7.00 and 60% chance of winning \$1.00	\$.30
5	50% chance of winning \$4.00 and 50% chance of winning \$3.50	50% chance of winning \$7.00 and 50% chance of winning \$1.00	-\$.25
6	60% chance of winning \$4.00 and 40% chance of winning \$3.50	60% chance of winning \$7.00 and 40% chance of winning \$1.00	-\$.80
7	70% chance of winning \$4.00 and 30% chance of winning \$3.50	70% chance of winning \$7.00 and 30% chance of winning \$1.00	-\$1.35
8	80% chance of winning \$4.00 and 20% chance of winning \$3.50	80% chance of winning \$7.00 and 20% chance of winning \$1.00	-\$1.90
9	90% chance of winning \$4.00 and 10% chance of winning \$3.50	90% chance of winning \$7.00 and 10% chance of winning \$1.00	-\$2.45
10	100% chance of winning \$4.00 and 0% chance of winning \$3.50	100% chance of winning \$7.00 and 0% chance of winning \$1.00	-\$3

Table 2: Calculated risk aversion within the HL lottery task.

Number of Safe Choices	**Range of Relative Risk Aversion U(x)=x^(1-r)/(1-r)**	**Risk Preference Classifications**	**Proportion of Choices for Treatment Group**	**Proportion of Choices for Control Group**
0-1	$r < -1.37$	Highly Risk Loving	8%	3%
2	$-1.37 < r < -.74$	Very Risk Loving	0%	0%
3	$-.74 < r < -.24$	Risk Loving	11%	13%
4	$-.24 < r < .19$	Risk Neutral	16%	7%
5	$.19 < r < .61$	Slightly Risk Averse	24%	23%
6	$.61 < r < 1.04$	Risk Averse	16%	43%
7	$1.04 < r < 1.54$	Very Risk Averse	9%	7%
8	$1.54 < r < 2.26$	Highly Risk Averse	5%	3%
9-10	$2.26 < r$	Stay in Bed	8%	0%

Table 3: Participants' demographics were anonymously collected and stored.

	Total (n=68)
Male	46%
Female	54%
Range Age	18-58
Average Age	24.75
Standard Deviation Age	9.87
Below Average Income	56%
Average Income	40%
Above Average Income	4%
High School Graduated	16%
Some College No Degree	50%
Bachelor's Degree	15%
Master's Degree	13%
Doctoral Degree	6%

Table 4: Correlation Matrix between the dependent variable and the independent variables, including a dummy variable for the treatment effect (n=68).

	Inconsistent	**Inv(ΔSCL)**	**Inv(ΔNS.SCRs)**	**ΔPA**	**ΔNA**	**CPT**
Inconsistent	1.0000					
Inv(ΔSCL)	0.1466 0.2328	1.0000				
Inv(ΔNS.SCRs)	0.1577 0.1990	0.6483* 0.0000	1.0000			
ΔPA	-0.0480 0.6975	-0.2042 0.0948	-0.0715 0.5621	1.0000		
ΔNA	-0.1826 0.1361	0.0216 0.8614	0.0331 0.1884	- 0.0700 0.5703	1.0000	
CPT	0.2991* 0.0132	-0.4765* 0.0000	-0.4569* 0.0001	0.0625 0.6125	-.1088 .3772	1.0000

Table 5: Correlation Matrix between the dependent variable and four control variables, including a dummy variable for the treatment effect (n=68).

	Inconsistent	**Age**	**Sex**	**Education**	**Income**	**CPT**
Inconsistent	1.0000					
Age	0.0541 0.6610	1.0000				
Sex	-0.1746 0.1544	0.0050 0.6561	1.0000			
Education	0.2151 0.0781	0.5722* 0.0000	0.0602 0.6257	1.0000		
Income	0.0735 0.5514	-0.1645 0.1800	0.0022 0.9855	-0.2340* 0.0548	1.0000	
CPT	0.2991* 0.0132	0.2070 0.0903	-0.2781* 0.0217	0.3208* 0.0077	-0.1244 0.3123	1.0000

Table 6: Three ordered logit models were analyzed using the dependent variable inconsistent. Standard errors are in the parentheses, $^{**}p < .01$, $^{*}p < .05$ two-tailed p-values test. Note, Inv(ΔSCL) was only used not Inv(ΔNS.SCRs) since there is a high correlation between the two independent variables.

Variables	**(1)**	**(2)**	**(3)**
Inv(ΔSCL)	4.73^{**} (.007)	4.83^{**} (.009)	5.19^{**} (.008)
CPT	2.14^{**} (.001)	2.09^{**} (.001)	2.13^{**} (.002)
ΔPA		.02 (.646)	.02 (.685)
ΔNA		-.12 (.098)	-.09 (.192)
Age			-.01 (.665)
Sex			-.35 (.529)
Observations	68	68	68
LR chi-squared	13.70	16.70	17.27
Prob > chi-squared	.001	.002	.008
Pseudo R-squared	.105	.129	.133
Mean VIF	1.29	1.18	1.25

Table 7. Average marginal effects after estimation of ordered logit models.

	# switches = 1	# switches = 2	# switches = 3
Equation (1)			
Inv(ΔSCL)	-.963** (.001)	.452** (.007)	.510** (.017)
CPT	-.432** (000)	.222** (.001)	.210** (.003)
Equation (2)			
Inv(ΔSCL)	-.941** (.002)	.430** (.008)	.510** (.017)
CPT	-.411** (000)	.208** (.003)	.203** (.003)
ΔPANAS pos	-.004 (.645)	.002 (.646)	.002 (.647)
ΔPANAS neg	.022 (.079)	-.010 (.104)	-.012 (.107)
Equation (3)			
Inv(ΔSCL)	-1** (.001)	.456** (.006)	.544** (.015)
CPT	-.415** (000)	.208** (.003)	.206** (.005)
ΔPANAS pos	-.003 (.684)	.002 (.686)	.001 (.686)
ΔPANAS neg	.018 (.177)	-.008 (.199)	-.010 (.199)
Age	.002 (.663)	0 (.663)	-.002 (.666)
Sex	.068 (.533)	-.031 (.545)	-.036 (.531)

Table 8: Three ordered logit models were analyzed using the dependent variable inconsistent. Standard errors are in the parentheses, $^{**}p < .01$, $^{*}p < .05$ two-tailed p-values test.

Variables	**(1)**	**(2)**	**(3)**
Inv(ΔNS.SCRs)	17.63^{**} (.007)	17.64^{**} (.009)	17.78^{**} (.009)
CPT	2.17^{**} (.001)	2.13^{**} (.001)	2.12^{**} (.002)
ΔPANAS pos		.04 (.406)	.04 (.408)
ΔPANAS neg		-.11 (.107)	-.103 (.178)
Age			-.004 (.844)
Sex			-.238 (.669)
Observations	68	68	68
LR chi-squared	15.25	18.43	18.64
Prob > chi-squared	000	.001	.004
Pseudo R-squared	.117	.141	.143
Mean VIF	1.26	1.14	1.20

Table 9. Average marginal effects after estimation of ordered logit models.

	# switches = 1	# switches = 2	# switches = 3
Equation (1)			
Inv(ΔNS.SCRs)	-3.512** (.001)	1.637** (.006)	1.875** (.017)
CPT	-.429** (000)	.216** (.001)	.212** (.003)
Equation (2)			
Inv(ΔNS.SCRs)	-3.352** (.002)	1.509** (.008)	1.843** (.019)
CPT	-.413** (000)	.206** (.002)	.206** (.003)
ΔPANAS pos	-.007 (.4)	.003 (.412)	.004 (.409)
ΔPANAS neg	.022 (.087)	-.009 (.113)	-.012 (.117)
Equation (3)			
Inv(ΔNS.SCRs)	-3.364** (.002)	1.513** (.008)	1.851** (.018)
CPT	-.410** (000)	.206** (.002)	.204** (.004)
ΔPANAS pos	-.007 (.402)	.003 (.414)	.004 (.410)
ΔPANAS neg	.019 (.161)	-.008 (.183)	-.010 (.187)
Age	0 (.884)	0 (.884)	0 (.884)
Sex	.045 (.667)	-.020 (.669)	-.024 (.669)

Chapter 3

Trick for a Treat: The Effect of Costume, Identity, and Peers on Norm Violations

Introduction

Clothing serves the social function of communicating information about the wearer to others. The economic importance of wearing the right clothes has led to memorable proverbs such as “clothes make the man”, “dress for success”, and “dress for the job you want and not the job you have”. While clothes clearly operate externally by communicating information to others, clothes may also have an internal effect by influencing one’s own sense of identity. Militaries dress their soldiers in uniforms as part of their socialization (Wakin, 2000; Akerlof and Kranton, 2005). Adam and Galinsky (2012) find that wearing a lab coat increases performance on attention-related tasks and school uniforms have been found to reduce disciplinary referrals (Sanchez et al., 2012). While the mechanisms for these effects are not entirely clear, the results suggest that clothes can affect people’s behaviors, perhaps through affecting one’s sense of identity.

We test this hypothesis by recruiting trick-or-treaters — children in costumed garb — during the American holiday of Halloween. We consider this a boundary condition for the effect of clothing on identity and behavior, as it is the day of the year in which participants are dressed to the greatest extremes. Moreover, trick-or-treaters often use their costume to take on the identity of specific characters from film or television, and these specific assumed identities may have particularly salient effects on behavior. In addition, the function of costumes for festivals and holidays, on its own, merits scientific study. Consumers spent $9 billion in 2018 on Halloween products (National Retail Federation, 2018) and many cultures around the world have developed traditions in which costume wear plays an important role. Interestingly, costume-wear often develops alongside traditions of norm-violations. Halloween evolved from the Celtic holiday of Samhain (Winkler and Winkler, 1970), in which costumes were used, in part, to hide one's identity during "tricks" or pranks (Miller et al., 1991). In Venice, the tradition of wearing masks during Carnival developed alongside activities that would otherwise be norm violations, such as mingling with other social classes, gambling, having clandestine affairs, reveling, and illicit activity (Walker, 1999; Burke, 2005).

Thus, it is historically and culturally apropos to measure the effect of Halloween costumes on ethical behavior. We use the lying game of Fischbacher and Föllmi-Heusi (2013) as our experimental paradigm. Trick-or-treaters privately roll a 6-sided die. If they report 1–5, they receive one candy, and if they report a 6, they receive the one candy and an additional bonus candy. If everyone tells the truth then the distribution of reported numbers would be uniform. Sixes reported in excess within a sample of individuals can be interpreted as evidence of lying for personal gain.

We manipulate three dimensions of the experimental conditions. First, we vary the stakes to price lying behavior. In the high-stakes condition, a reported six earns two bonus candies (three total) instead of the one bonus candy. Second, we vary the beneficiary of the lie to test whether lying for others is normative. On the one hand, incurring a psychic cost of lying to benefit someone else may be unappealing for some, but on the other hand, violating a norm in order to benefit someone else may be itself perceived as normative as there is no possible intimation of selfish behavior. In the baseline condition, the beneficiary is one's self. In the "other" condition, reporting a six earns someone in the next group of trick-or-treaters an additional candy. In the "both" condition, the trick-or-treater rolls two dice, the first affects one's own payoff, and the second affects the payoff of a trick-or-treater in the next group. The both condition allows us to measure whether behavior in the "self" ("other") condition spills over to behavior in the "other" ("self") condition.

Third, we vary the degree of salience of one's costume to observe how it affects the trick-or-treater's ethical behavior. The variation in Halloween costumes leads to a natural separation in costumes between heroes and objects of admiration on the one hand, and villains and creatures of a wicked nature on the other. For example, the most popular costumes in our sample are (in order of popularity) a unicorn, Spiderman, Batman, Master Chief of the video game *Halo*, evil clown, vampire, Jason from *Friday the 13th*. The first four would be considered admirable by most people, while the latter three would be considered wicked by most people. In the treatment condition, we ask the trick-or-treater who they are dressed as, whether that character is a "good guy or bad guy", and whether that person does "good things or bad things". This is intended to draw salience to the person's costume and the character's ethical orientation. In the control condition, we ask the same questions but after the participant has already reported their dice rolls.

Additionally, natural variation in participant age and arrival affords us the ability to answer two other thematic questions. (1) Does age affect lying? This is an important question, as a vast developmental psychology literature shows that children's cognitive abilities and behaviors mature at specific ages. (2) To what extent is lying in children influenced by peers?

The experiment was run at a house on a street that hosts large crowds of trick-or-treaters. Participants lined up and were allowed to advance to the porch in groups of 10, where upon they were told the rules of the game. While participants advanced to the door of the house one-by-one, participants behind them could likely hear their reports. Because we recorded the order in which participants lined up, we can measure the effect of reporting a six on subsequent participants' reports. Specifically, we test the extent to which trick-or-treaters are affected by same gender versus opposite gender peers. These two questions fall within the paper's broader theme of identity, in this case relating specifically to the trick-or-treater's age and gender identities.

We found frequent occurrence of six with 40% of participants reporting a six in the baseline condition, which is approximately 23 percentage points more sixes than by chance alone. Stakes had no effect on the number of reported sixes. The other-condition strongly reduced the frequency of reporting a six. Many trick-or-treaters did not view the cost of lying to be worth helping out an anonymous stranger. In the both-condition, the occurrence of reporting a six to benefit one's self decreased, while the occurrence of reporting a six to benefit other's did not change. The result suggests that participants' honesty in reporting for others spilled over to reporting for self but not vice versa.

Next, we test the effect of costume salience on participant's behavior. Costume choice is endogenous: the correlation between costume and lying would not necessarily indicate a causal relationship. However, increasing the salience of a costume is random and it is hypothesized to have different effects for those who are dressed as characters of admiration (heroes or creatures of beauty), and those who are dressed as wicked characters.

We hypothesized that drawing attention to the ethical orientation of one's assumed identity would lead to behavior consistent with that identity. That is "good guys" whose costumes are rendered salient would lie less than "good guys" whose costumes are not made salient and "bad guys" whose costumes are rendered salient would lie more than "bad guys" whose costumes are not made salient. In fact, we found the opposite. The salience condition caused good guys to lie more, by about 12 percentage points, than good guys in the control condition.

Bad guys with the salience of their costumes lied significantly less, by about 27 percentage points less, than bad guys in the control condition. We offer two possible interpretations of these results. The results are consistent with a moral licensing effect for good guys (Secilmis, 2018; Lasarov and Hoffmann, 2018). Making "good guy" salient may have made individuals feel more justified in committing a norm violation, and the reverse effect may have operated on "bad guys". Alternatively, the effect may be driven by a feeling of being monitored. The trick-or-treaters declared to an observing adult whether they were a "good guy" or a "bad guy". This may have changed the trick-or-treaters' perceptions about the extent to which they were being monitored. Perhaps self-declared "good guys" felt as though they were putting themselves in our good graces, while self-declaring oneself as a "bad guy" felt like it would warrant greater monitoring from us. We cannot test between these two hypotheses but the physical structure of our experiment casts some doubt on the plausibility of the latter hypothesis.

Finally, examining the natural variation of our sample, we find an inverted-U pattern for the effect of age, with lying peaking at age 12. We find large and statistically significant peer effects. One additional person reporting a six out of a group of five participants increases the probability of reporting a six by about 5 percentage points. Interestingly, as we decompose this effect by gender, we find that most of the effect operates within gender. That is girls follow girls and boys follow boys. It appears that gender identity prominently moderates the peer effect.

We view our main contribution as showing that salience of one’s clothed identity affects ethical behavior in the direction of a moral licensing/self-conscious effect. Additionally, we find that gender identity influences behavior, with girls emulating girls, and boys emulating boys in their reporting behavior. We also provide evidence on whether lying for others is viewed as normative for children, and provide evidence for the existence of within-person spillover effects and age effects.

Literature Review

Fischbacher and Föllmi-Heusi (2013) developed an ingenious experimental paradigm for the measurement of aggregate lying at the population level. In the original study participants roll a die privately and obtain a reward based on the result they report to the experimenter. They found that about 20% of participants lie to the fullest extent possible while 39% of them are fully honest. The experiment has generated numerous variants (see Abeler, Nosenzo and Raymond, 2019, for a meta-analysis). Notably, Cadsby, Du and Song (2016) run a variant in which they find that people lie to increase the payoff of an in-group member even though such a lie does not affect their own monetary payoff.

A few recent economic papers have examined lying in children. Bucciol and Piovesan (2011) conduct a similar experiment in a children's summer camp with ages 5-15. They find no association between lying and age, however they only report a linear specification and their sample is small in their baseline condition (N=81). Glätzle-Rützler and Lergetporer (2015) find that 10 and 11 year olds are more likely to lie than 15 and 16 year olds, consistent with our age pattern. Maggian and Villeval (2016) run a somewhat different game in which lying can yield higher payoffs, with 7–14 year olds. They find that 9-10 year olds lie more than 7-8 year olds or 11-14 year olds. This is similar to our finding as we also have an inverted-U as a function of age, though our peak age is 12. Brocas and Carrillo (2019) find that middle-schoolers aged 11-14 lie significantly more than the other age groups (5-8 year olds, 8-11 year olds, 14-17 year olds, and undergrads), which confirms our results. They additionally provide evidence on lying as a function of age when the benefactor is another child. They find that only middle-schoolers exhibit lying in the aggregate, and they lie to reduce the other child's payoff! In aggregate, we find that our subjects lie to benefit others;

the difference between these results may be due to design differences and differences in payoffs (for example, in their experiment, payment is strictly increasing in the die roll). Finally, on the topic of stakes Fischbacher and Föllmi-Heusi (2013) and Mazar, Amir and Ariely (2008) do not find an effect of stakes on lying.

The research on peer-effects of norm-violations consistently finds that violations are contagious. Gino, Ayal and Ariely (2009) conduct a cheating experiment in which a confederate publicly announces that he has completed a task in an impossibly short amount of time. If the confederate is an in-group member, cheating goes up relative to the baseline. Interestingly, if the confederate is an out-group member (dressed in the clothes of a rival school) cheating goes down relative to the baseline. Diekmann, Przepiorka and Rauhut (2015) and Bicchieri, Dimant and Sonderegger (2020a) include conditions in lying experiments in which subjects are informed about the lying behavior of others. They find that information about the extent of lying in the population increases lying compared to the control condition. In a charitable contribution game, Bicchieri et al. (2020b) find similarly that learning about the empirical distribution of other's contributions causes less compliance with giving. Our results strengthen this literature, showing that direct observation of peers reporting a six increases reporting a six. Our results extend this literature by showing that contagion occurs within gender rather than across gender, in a sample of children. Consistent with the findings of Gino, Ayal and Ariely (2009), we show that the identity of the observed cheater moderates the contagion effect. They show it occurs via school affiliation, we show that the results extend to gender identity.

Abeler, Nosenzo and Raymond (2019) summarize the theory literature on the that

Fischbacher and Föllmi-Heusi (2013) paradigm. They categorize models into those that have only, those that have lying costs with a preference for conformity, and those that have lying costs and a preference for honest reputation. They conduct a Herculean meta-analysis of the experimental literature as well as an additional large experiment of their own and they find that only the last category of models, lying costs and a preference for an honest reputation, explains the pattern of results. Our results are consistent with this finding in the sense that (1) lying in our study is not maximal and this can be explained by lying costs. (2) Our salience effect — salience causes good guys to lie more and bad guys to lie less — could be recast as a preference for an honest reputation. The salience may make good guys feel like their ex-ante reputational capital is higher, and the opposite may be true for bad guys. If there are diminishing returns to reputation, then those who feel they have higher reputation are more likely to spend it for a sweeter physical reward. This could generate our result. Of course, we don't observe or directly manipulate self-perceived reputation in our study, so this latter point is speculative. We didn't find economic models that relate to our other research questions.

To the best of our knowledge, no study has examined the effect of the salience of clothing on lying behavior. Research on the economics of clothing is limited but it has a long history going at least as far back as Veblen (1899). Veblen observed that restrictive or delicate clothing can be a costly signal indicating that the wearer is of the well-to-do “leisure class”, since a laborer would be unable to function in such clothes. Clothing also communicates people’s roles. For example, a concerned citizen can identify a police officer during an emergency thanks to the officer’s uniform. Given that clothes communicate a person’s identity to others, researchers have posited that clothes may also affect one’s self-identity. Adam and Galinsky (2012) find that wearing a lab coat increases performance on attention-related tasks. School uniforms can reduce discipline referrals by 9.7% (Sanchez et al., 2012).

Civile and Obhi (2017) randomly assign participants to wear police-style uniforms and then have them engage in an attention-related task. They find that uniforms cause participants to attend more to images associated with lower socio-economic status. The results establish that wearing clothes associated with particular roles in society, or identities, can causally affect behavior.

Wearing costumes on Halloween that assume the identity of specific characters has become a mainstream practice in American culture. A lab coat or a police officer’s uniform have metonymic relationships with the institutions they represent. In contrast, many Halloween costumes reflect specific characters (e.g. Batman or Moana) as opposed to broader roles in society. Thus it is not clear whether the greater specificity common in Halloween costumes can enable the kind of identification and change in behavior exhibited in the aforementioned studies. However, casual observation suggests that at least some level that they like to act out in pretend play. Pretend play often accompanies the wearing of costumes: Reys and Lukes swing their lightsabers against imaginary stormtroopers, Moanas steer their imaginary ships to safety, and vampires suck the imaginary blood of their real siblings (much to their parents' dismay). Indeed, it seems that an important reason why Halloween is so attractive to children is that it facilitates enjoyable roleplay.

Halloween evolved from the Celtic holiday of Samhain during which it was believed that spirits and souls of the dead would return to earth (Sterba, 1948). The festivities involved people going door-to-door in costume reciting verses in exchange for food (Linton and Linton, 1950; Ward, 1981). Later in the 16th century Scotland, revelers would wear masks or painted faces threatening to do mischief if they were not given food (Linton and Linton, 1950; Belk, 1990). Costume wearing and norm violations emerged contemporaneously in this context, and the co-emergence of the two institutions have arisen in other cultures as well. The wearing of costumes in Venice for Carnival originated alongside traditions of mischief-making and intermingling of social classes that were otherwise discouraged from mixing (Feil, 1998). We note several social functions that these costumes may have served.

First, disguising one's outward identity while begging may have been a way for those individuals to avoid harm to their reputation and the associated shame. Second, disguising one's outward identity while violating any rule, be it legal or an implicit cultural norm, has the obvious benefit of avoiding punishment. Thus, disguises are complements with norm violations, the presence of one in a tradition increases the marginal value of including the other.

Masking one's identity may make one more prone to norm violation. In addition, research has shown that rendering specific aspects of one's identity salient can change one's propensity to violate norms. Recent studies find that bank employees become dishonest and clergymen become more honest when their professional identity is rendered salient (Celse and Chang, 2017). Cohn, Maréchal and Noll (2015) show that increasing the salience of prisoner's criminal identity increases dishonest behavior. In these studies, making an aspect of the individual's identity salient promotes ethical behavior congruent with that aspect. However, a separate thread of research has established a potential countervailing force.

Moral licensing is the psychological phenomena in which boosts to self-image increase engagement in unethical behavior (Nisan and Horenczyk, 1990). Sachdeva, Iliev and Medin (2009) had participants write a short story about themselves or someone they knew using morally positive trait words (e.g., fair, kind) or morally negative trait words (e.g., selfish, mean). Participants assigned to write about themselves using positive traits donated the least out of the four conditions and those who wrote about themselves using negative traits donated the most. Khan and Dhar (2006) obtained similar results, finding that participants asked to imagine helping others donated less to charity than control subjects, and Mazar and Zhong (2010) show that people act less altruistically and are more likely to cheat and steal after purchasing green products than after purchasing conventional products. Jordan, Mullen and Murnighan (2011) had participants recall one's own moral or immoral past actions or another's moral or immoral past actions. They find that people who recalled their own immoral behavior reported greater participation in moral activities, reported stronger prosocial intentions, and showed less cheating than people who recalled their own moral behavior. Similarly, Clot, Grolleau and Ibanez (2014) find that participants who recalled their own moral actions subsequently cheated more to get a higher payoff than participants who did not recall their moral actions. We extend the literature by measuring whether drawing salience to one’s costume affects the propensity to lie, differentially for those dressed as “good guys” vs. “bad guys”.

Methods

1.1 Sample and Setting

Our experiment was conducted on the dark and stormy night of Halloween, October 31st, 2018, at the house of one of the authors in a suburb of Los Angeles. The neighborhood is a destination for trick-or-treating amongst nearby communities. A typical home in this neighborhood is visited by about 1,000 trick-or-treaters. Trick-or-treaters who approached the house between 6:00 and 9:30pm participated in the experiment. They were told that they could play a game to win candies. In total, 544 trick-or-treaters participated. Other participants of a spiritual and malevolent nature, may have participated, undimensioned and unseen, without our knowledge.

1.2 Experimental Procedure

The experiment proceeded as follows. An experimenter advertised to passing trick-or-treaters that they could play a game to win candies. Trick-or-treaters queued in front of the porch and were randomly assigned to two lines. One line led to the no-salience condition and the other line led to the costume salience condition. All subjects were given an ID card with a number (1-10) and were instructed that they would need their ID card to exchange for candy. All subjects were asked their age. A quietly observing experimenter recorded subjects' answers discretely along with other information such as gender, whether parents accompanied them or not, time of the day, and the specific ID number. Those in the salience condition were asked additional questions, "Who are you today?", "Is (answer to the previous question) a good guy or a bad guy?" and "Does (answer to the first question) do good things or evil things?” The subjects in the no-salience condition were asked the same questions but only at the very end of the experiment. Subjects advanced approximately in the order they arrived.

We use the lying game introduced by Fischbacher and Föllmi-Heusi (2013). Ten subjects were allowed to the porch at a time at which point we explained the rules of the game: they would role a 6-sided die in a paper cup. If they rolled a 1-5, they would get one candy; if they rolled a 6, they would win bonus candy. We also stated very clearly, “You don’t need to show us the dice. Just tell us the number.” Experimenters at the end of the line asked for the number and gave the promised number of candies.

1.3 Experimental Treatments

The experiment has $2 \times 3 \times 2$ conditions: we varied the stakes (high vs. low), we varied who the beneficiary of the bonus candy was (self vs. other vs. both), and we varied the salience of subjects’ Halloween costumes (no-salience vs. salience). As mentioned above, the salience condition was randomized at the individual level as trick-or-treaters approached the house. The stakes and beneficiaries were randomized by group. We cycled through the six conditions, alternating after every group of 10 subjects. We had a total of 55 groups.

We varied the stakes in order to price the effects of the other treatments. In both conditions, subjects receive one candy for reporting any number. In the low-stakes condition, they earned one bonus candy for reporting a six, and in the high stakes condition they earned two bonus candies for reporting a six. Though the stakes are low, a pilot study conducted the year prior suggested they would be adequate. In the pilot, the low-stakes condition paid 1 candy for a reported one through four, 2 candies for a reported five, and 3 candies for a reported six. We found an excess of fives and an even greater excess of sixes. In the high-stakes condition the rewards were 1 candy, 3 candies, and 5 candies. In that study we observed a significant effect of high stakes. We expected that the 1 additional bonus candy in the high-stakes condition would be sufficient.

In the “self” condition, the subject was the recipient of the bonus candy, and in the “other” condition the recipient of the bonus candy was the subject in the next group with the same ID number. In the “both” condition, subjects rolled one die for themselves and one die for a subject in the next group. As mentioned above, subjects in the salience condition were asked questions prior to the instructions, while subjects in no-salience condition were asked the same questions after the number was reported.

1.4 Additional Covariates

In our regressions we include additional covariates. We wish to test whether there is an age pattern in the propensity to lie. We censor ages from 4 to 19, to deal with outliers (there are the rare 50 year-old trick or treaters). We estimate a quadratic specification for age, allowing for curvature and a slope sign change.

Subjects were admitted to the porch in groups of 10 at which point they were handed dice and explained the instructions. The children were in close proximity to each other as they lined up to report their number. It would have been easy for children in line to overhear the reports and earnings of children in front of them. This accidental feature of the design allows us to identify peer effects. The exogenous arrival rate of trick-or-treaters creates variation in the composition of each group and our assignment of ID number creates variation in the sequential order of the reporting. We estimate a coefficient for a "peer proportion" variable defined as the proportion of previous participants in the group that reported a six. If the person is first in their group we define peer proportion as zero. We decompose the peer-proportion effect by gender. We define "female peer proportion" as the proportion of previous female participants in the group that reported a six, and define the "male peer proportion" variable in the analogous way for males. If there were no previous females in the group then "female peer proportion" is defined as 0 and if there were no previous males in the group then "male peer proportion" is defined as zero.

We include several covariates as controls to increase statistical power. In all statistical models in which age is not the primary focus the quadratic age variables are not included; we instead use a more flexible specification using 11 age categories. Ages 4 and less are the first category, each age from 5 to 11 are their own categories, and ages 12-14, 15-18, and 19+ are the last three categories. Some children are accompanied by their parents. Obviously, the presence of one's parents could have an effect on a child's propensity to lie so we include a parent indicator as a control. We also include a gender indicator as a control variable in all models.

1.5 Multiple-Hypotesis Testing

A potential weakness of our research design is the large number of variables of interest. Naïve multiple-hypothesis testing can inflate the chance of having statistically significant results by chance, even under the null hypothesis. To control for false positives we include a false discovery rate (FDR) analysis. We use the method of Benjamini, Krieger and Yekutieli (2006), which is the sharpened two-step FDR procedure. This procedure produces less false negatives than the original FDR procedure. Any coefficient displayed in a table is included in the set of hypothesis tests we use for our paper-wide FDR correction. The q-values are equivalent to p-values that have been adjusted for multiple hypothesis testing.

Results

1.6 Summary Statistics

Table 1 shows the summary statistics of our sample. Among a total of 544 subjects who participated in our experiment, about 53% are female, and 41% are accompanied by parents. The ages of our subjects range from 0 (baby) to 50 years old (as some parents also joined the game) with an average age of 9.46 years old. The very young children were accompanied by their parents. In the analysis, we bottom-code participants age at 4 years-old and top-code participant age at 19 years-old due to the small samples outside of this range. About 24% wore a self-reported bad-guy costume. The most popular costumes in order of popularity were: unicorn, Spiderman, Batman, the *Halo* video game main character, clown, vampire, and Jason from *Nightmare on Elm Street*.

The frequency of reporting a six within our sample is 41%, which is significantly above the probability of rolling a six at 16.7%. This implies that a substantial number of our subjects lied in the experiment. While traditional economic theory predicts that when there is no cost to lying, people would always report a six, our results show that at least some subjects are honest. The results are consistent with Fischbacher and Föllmi-Heusi (2013), with a substantial degree of lying.

Table 2 displays the raw totals and percentages of reporting a six by condition. The high stakes and low stakes conditions both have approximately 40% reporting a six. The differences between the beneficiary conditions is sizable. While 54% report a six in the self condition, only 39% report a six in the other condition. Reporting a six further drops in the both condition with 39% reporting a six for themselves and only 27% reporting a six for others. Reporting a six is just as likely in the no-salience and salience conditions with only a

Table 1: Summary Statistics

	mean	sd	min	max
Six	0.41	0.49	0	1
Six_Self	0.47	0.50	0	1
Six_Others	0.33	0.47	0	1
Group	27.99	15.72	1	55
Age	9.46	5.38	0	50
Parents	0.41	0.49	0	1
Self Condition	4.61	1.66	1	6
Others Condition	4.28	1.66	1	6
Female	0.53	0.50	0	1
High Stakes	0.51	0.50	0	1
Salience	0.50	0.50	0	1
Bad Guy	0.24	0.43	0	1
Good and Bad	0.02	0.13	0	1
Age Category	9.33	4.11	4	19
Six within Group	2.24	2.25	0	12
Six within Group_Female	1.14	1.46	0	10
Six within Group_Male	1.10	1.34	0	8
Observations	544			

1 percentage point separating the two. If there is a gender effect, it appears small, with boys reporting a six only 2 percentage points more than girls. True to their names, there does appear to be a difference between good guys and bad guys with six reports at 39% vs. 43%. To better assess the effects of the treatments we turn to regression analysis.

Table 2: Raw Probabilities

	Outcome				
	Not Six		**Six**		**Total**
	No.	%	No.	%	No.
Stakes Low stakes	202	60	137	40	339
High stakes	211	60	142	40	353
Total	413	60	279	40	692
Beneficiaries					
Self	84	46	100	54	184
Other	104	61	67	39	171
Both	225	67	112	33	337
Both - Self	103	61	66	39	169
Both - Other	122	73	46	27	168
Total	413	60	279	40	692
Salience No Salience	203	59	140	41	343
Salience	210	60	139	40	349
Total	413	60	279	40	692
Gender					
Male	191	59	133	41	324
Female	220	61	141	39	361
Total	411	60	274	40	685
Costume					
Good Guy	279	61	181	39	460
Bad Guy	83	57	62	43	145
Total	362	60	243	40	605

1.7 Main Results

We display the analysis of our treatments in Table 3. All columns report average marginal effects of a logistic regression. Each cell of the table contains four statistics. The number in the upper-left corner is the average marginal effect of the variable, the number in the lower-left corner in parentheses is the standard error. The number in the upper-right corner in italics is the naïve p-value of the average marginal effect and the number in the lower-right corner is the sharpened FDR q-value in square brackets. The q-value can be interpreted as a p-value that corrects for multiple hypotheses. Each cell of a table is considered a hypothesis test for the purposes of the paper-wide FDR corrected q-values.

Column (1) regresses an indicator for reporting a six as the outcome on the stakes condition, beneficiary conditions, gender, and the presence of parents. Additional unreported controls include age categories. The effect of the high-stakes condition is not significant. The "other" condition significantly reduces the frequency of reporting a six relative to the baseline "self" condition. Subjects are less willing to lie to benefit an anonymous stranger than they are if the benefit accrues to themselves. The negative coefficient in the "both" condition indicates that having two die rolls reduced the occurrence of reporting a six to benefit one's self. The results show no effect of gender. In Column (2), we no longer control for age using categorical variables but instead include a

Table 3: Reporting a Six – Logistic Regression with Average Marginal Effects

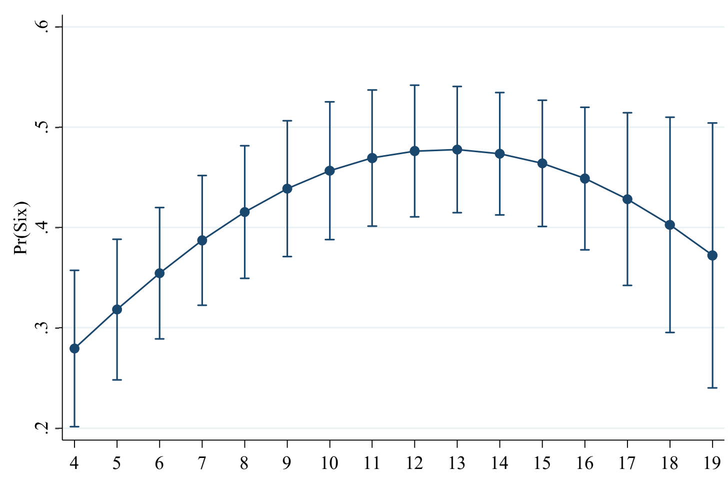
	(1)		(2)		(3)	(4)
	-0.012	*0.809*	-0.002	*0.958*	-0.001	*0.981*
	-0.137**	*0.018*	-0.115**	*0.015*	-0.117**	*0.013*
	-0.14**	*0.046*	-0.092	*0.109*	-0.094	*0.107*
	0.018	*0.828*	-0.016	*0.832*	-0.017	*0.827*
Female					0.008	*0.885*
					(0.052)	[0.657]
Age	0.067***	*0.002*				
	(0.021)	[0.039]				
Age^2	-0.003***	*0.009*				
	(0.001)	[0.065]				
Peer Proportion			0.268***	*0.0*		
			(0.059)	[0.001]		
Female Peer Proportion					0.277	*0.102*
					(0.17)	[0.125]
Female x Female Peer Proportion					0.036	*0.865*
					(0.213)	[0.657]
Male Peer Proportion					0.285***	*0.004*
					(0.1)	[0.046]
Female x Male Peer Proportion					-0.104	*0.461*
					(0.142)	[0.383]
Female Peer Proportion on Female					0.314***	*0.007*
					(0.117)	[0.06]
Male Peer Proportion on Female					0.18*	*0.08*
					(0.103)	[0.125]

Notes: The dependent variable is whether the subject reported a six or not. All models are logistic regression with age categories, gender and parents included as controls. The first column in each cell reports average marginal effects with standard errors in parentheses. The second column in each cell reports naive p-values in italics and FDR adjusted q-values in brackets. Standard errors are clustered by group. Each cluster is a set of 10 subjects who were given instructions at the same time. Subjects in the "Both" treatment reported two outcomes. To analyze peer effects in Column (3) and Column (4), we exclude the first subject from each group whose probability of reporting a six should not be affected by anyone else. *** p<0.01, ** p<0.05, * p<0.1.

second-order polynomial in order to fit an age trend. The coefficient on age is positive and significant while the coefficient on age is negative, indicating that lying exhibits an inverted U pattern. Figure 1 displays the predicted six-reports as a function of age with 95% confidence intervals. Note that the probability of reporting a six is significantly higher than 20% across the whole domain, which is above 16.7%, the expected frequency of reporting a six under truth-telling. These results indicate that lying happens at all age levels. The probability of reporting a six peaks at age 12 and subsequently decreases thereafter.

Figure 1: Reporting a Six as a Function of Age

Predictive Margins with 95% CIs



Notes: Probability of reporting a six across ages. Error bars indicate 95% confidence intervals. Ages are censored at 4 and 19.

In Column (3) we test for peer effects on reporting a six. Recall, the "Peer Proportion" variable as the proportion of previous reports within the group that were six. The coefficient is positive and statistically significant with a p-value < 0.001. This provides evidence that the probability of reporting a six is influenced by previous subjects in the group reporting a six. If one additional child out of five previous children reports a six, it leads to a 5.36 percentage point increase in the probability of a subsequent child reporting a six.

In Column (4) we decompose the peer effect by gender. "Female Peer Proportion" measures the proportion of female subjects before the subject in question who reported a six, and "Male Peer Proportion" does likewise for male subjects. While gender does not have a direct effect on one's lying behavior, our data show a difference in how boys and girls are influenced by the other subjects in their group. The coefficient on "Female Peer Proportion" is the effect of previous female subjects on male subjects and it is positive but not significant. The coefficient of "Male Peer Proportion" is the effect of previous male subjects on male subjects, which is large, positive, and statistically significant ($p = 0.004$). As we turn to girls, the coefficient on "Female Peer Proportion" plus the coefficient on "Female × Female Peer Proportion", 0.314 is positive, and significant ($p = 0.007$). The effect of boys on girls is "Male Peer Proportion" plus "Female × Male Peer Proportion" which is 0.18, and marginally significant at p=0.08. To summarize, boys emulate boys and girls emulate girls. We interpret this as evidence that children take their social cues within gender, tending to emulate others who share their gender identity.

In Table 4 we turn to the effects of the salience of subjects' costumed identities. Column (1) contains the same regressions as from Table 3 except we add the additional indicator variables for self-reported “bad guy”, salience condition, and the interaction between the two. In this table we drop subjects who self reported as “both a good guy and bad guy” or as “neither a good guy or bad guy”. We hypothesized that subjects would behave more congruously with their costumed character in the costume salience condition. Because “good guys” and “bad guys” are expected to respond in opposite directions, the interaction term is essential. Instead, we find evidence for the opposite effect. The coefficient on salience is the effect on “good guys” and it is positive but insignificant. The coefficient on the interaction term “Salience × Bad Guy” is the differential effect of costume salience on “bad guys” relative to “good guys” it is negative and significant at $p = 0.036$. The results are tantalizing yet unconvincing. We hypothesized that the relatively lower rates of lying in the “other” conditions may be attenuating the effect.

In Column (2), we decompose the effect of costume salience into “self” and “other” conditions. We include interactions between salience and “other”, “bad guy” and “other”, and the triple interaction between salience, “bad guy”, and “other”. The coefficient on “bad guy”, which is positive and significant at $p = 0.019$, implies that those wearing “bad guy” costumes are about 21 percentage points more likely to lie to benefit themselves than those wearing a “good guy” costume. Costumes are self-selected so this could be either a selection effect, a causal effect of wearing the costume, or some combination thereof. The manipulation is not the costume but the salience of the costume. The coefficient on salience is interpreted as the effect of the costume salience condition on “good guys” when the beneficiary is one’s self. The coefficient increased relative to Column (1), and is now significant at $p = 0.049$. Interestingly, the coefficient on “Salience × Bad Guy”, which is the differential effect of salience on “bad guys” relative to “good guys” when one’s self is the beneficiary, is now larger in magnitude, negative and significant at $p = 0.003$. The overall effect of costume salience on “bad guys” when one’s self is the beneficiary, is the sum of the coefficients on salience and “Salience × Bad Guy”, which is -0.153 negative and marginally significant at the $p = 0.08$.

The fact that the coefficients on “Salience × Other Treatment” and “Salience × Other Treatment × Bad Guy” have opposite signs relative to their self-condition counterparts (i.e. “Salience” and “Salience × Bad Guy”) confirms that the salience of costume had an effect primarily in the “self” condition. The salience of costume had no significant effect in the “other” condition.

Table 4: The Effect of Priming on Reporting a Six – Logistic Regression with Average Marginal Effects

	(1)		(2)	
High Stakes	-0.044	*0.358*	-0.046	*0.333*
	(0.048)	[0.299]	(0.047)	[0.285]
Other	-0.122**	*0.044*	0.001	*0.99*
	(0.06)	[0.093]	(0.084)	[0.657]
Both	-0.138**	*0.048*	-0.14**	*0.044*
	(0.07)	[0.093]	(0.07)	[0.093]
Other x Both	0.001	*0.99*	0.006	*0.943*
	(0.09)	[0.657]	(0.088)	[0.657]
Salience	0.036	*0.435*	0.115**	*0.049*
	(0.046)	[0.37]	(0.058)	[0.093]
Bad Guy	0.089	*0.24*	0.214**	*0.019*
	(0.076)	[0.239]	(0.092)	[0.078]
Salience x Bad Guy	-0.169**	*0.036*	-0.268***	*0.003*
	(0.08)	[0.093]	(0.089)	[0.039]
Salience x Other			-0.168**	*0.044*
			(0.083)	[0.093]
Bad Guy x Other			-0.239**	*0.03*
			(0.11)	[0.092]
Salience x Bad Guy x Other			0.304	*0.112*
			(0.191)	[0.125]
N	600		600	
Clusters	55		55	

Notes: The dependent variable is whether the subject reported a six or not. All models are logistic regression with age categories, gender and parents included as controls. The first column in each cell reports average marginal effects with standard errors in parentheses. The second column in each cell reports naive p-values in italics and FDR adjusted q-values in brackets. Standard errors are clustered by group. Each cluster is a set of 10 subjects who were given instructions at the same time. Subjects in the "Both" treatment reported two outcomes. We managed to record the costume information for only 464 subjects. *** p<0.01, ** p<0.05, * p<0.1.

Drawing attention to “good guys” costumed identity causes them to lie more, while drawing attention to “bad guys” costumed identity causes them to lie less. This result is opposite to what we predicted.

1.8 Power Analysis

A potential concern of this analysis is that statistical tests may not be sufficiently well powered. We have 540 subjects placed into 2 × 3 × 2 randomly assigned conditions. In addition, we test non-assigned features like age and peer effects. In many of our tests, statistical power would not appear to be problematic as the tests pool across conditions, and appropriately so as the other conditions and variables of interest are orthogonal. However, for some analyses we estimate triple interactions leading to splits in the range of 1/8th of the sample (about 75 subjects) per cell. If statistical power is low, there is a risk that statistically significant results are lucky, generated from sampling variation and not from a true underlying effect.

We conduct an ex-post power analysis using Monte Carlo methods. We subject all of our results that are significant at the $p < 0.1$ level to this analysis. For each given logit regression we predict the probability that an individual will report a six. We then generate 1,000 data sets, each randomly assigning the outcome (reporting a six) to an observation based on the predicted probability. On each of these data sets we run the same regression and estimate the average marginal effects. The proportion of data sets that generated a significant coefficient at $p < 0.05$ is our estimate of statistical power for that coefficient. We conduct this analysis for every regression specification in the paper (four in Table 1 and two in Table 2; six total).

Conclusion and Discussion

We found that our high-stakes manipulation had no effect. This is in-line with the results of previous studies (Mazar, Amir and Ariely, 2008; Fischbacher and Föllmi-Heusi, 2013). However, we ran a pilot in the preceding year with a similar design and slightly different stakes and found an effect. We view the result in this paper as a failed manipulation as we suspect that our high stakes were not sufficiently high. We believe that we cannot conclude much from this manipulation.

Changing the beneficiary had a large impact on lying. The “other” condition had much less lying than the “self” condition. Gino and Pierce (2010) and Cadsby, Du and Song (2016) show that people lie for others when they care for them, especially if the other person is part of their group. Michailidou and Rotondi (2019) find that individuals are not willing to lie to benefit an out-group member and that lying for an in-group member is uncommon. Our results are consistent with these findings as the beneficiary in our “other” condition is an anonymous stranger. We also find that the “both” condition reduces lying for one’s own gain but not for the gain of others. Wiltermuth (2011) and Gino, Ayal and Ariely (2013) find that people are more likely to view dishonesty as morally acceptable when their dishonesty would benefit others. The past studies found that children lie more for both themselves and others when there is an opportunity to lie to help others. Our results are contrary to this finding. Instead, honesty for reporting the “other” die spilled over to reporting the “self” die but not vice versa. This could be caused for a number of reasons. The cognitive load placed on the children for engaging in the two activities may have caused them to lie less. Or the inconsistency between lying about one report but not the other may have made rationalizing lying more difficult. A third potential explanation is that the children may have thought they would be more likely caught lying if they reported a six for themselves and not a six for "other". We cannot test between these different hypotheses unfortunately.

We find no direct effect of gender. This is consistent with the result of Nieken and Dato (2016). However, some research suggests that men are more likely to lie to advance themselves, while women are more likely to lie to advance others (DePaulo et al., 1996; Feldman et al., 2002; Erat and Gneezy, 2012; Dreber and Johannesson, 2008). The sample here differs from other studies as it is primarily children.

It is surprising to us that costume salience led to behavior incongruous with one's costume. Past research found the salience of identity to have effects congruous with the identity (Cohn, Fehr and Maréchal, 2014; Cohn, Maréchal and Noll, 2015; Celse and Chang, 2017). We offer some speculation regarding this finding. One possibility is that the costume led to congruous ethical behavior earlier in the day and this bolstered a moral licensing/self-conscious effect later during our experiment.

However, for this explanation to make sense it would require that the salience manipulation not only made the ethical orientation of one's costume more salient but also made past recent ethical actions more salient. While possible, the manipulation made no mention of recent activity.

An alternative explanation is that the salience manipulation felt like an announcement or confession of moral disposition to an *observing* adult. Past research has shown that people conform more to social norms when they feel like they are being observed even though they are not actually being observed. For example, Haley and Fessler (2005) found that the image of stylized eyespots increased giving in a dictator game. Mol, van der Heijden and Potters (2020) showed that subjects in a virtual reality environment were less likely to cheat when a virtual observer was watching, compared to when the virtual observer was looking at a smartphone. If the children thought that the adults were monitoring them, publicly announcing to an adult that they are “good guys” may have reduced the feeling of being observed, while announcing to an adult that they are “bad guys” may have increased the feeling of being observed.

There are a couple of reasons we believe this explanation is less convincing than the moral licensing explanation. First, several papers find no effect of watching eye cues (Pfattheicher, Schindler and Nockur, 2019; Ayal, Celse and Hochman, 2019), suggesting that the effect may not be very robust. Second, and more importantly, the experimenters who asked the priming question were different from the experimenters who ran the game and gave the candy. The experimenters who asked the priming question were also physically distant, at about 25 feet away from the point at which the trick-or-treaters reported their die rolls. This means that for this “feeling of differential monitoring” hypothesis to be true, trick-or-treaters would have to feel that their communication to one experimenter affected other experimenters 25 feet away without the use of communication. This sounds less plausible to us but it remains a possibility. An alternative design that makes the salience manipulation private could potentially disentangle this effect.

While modern practice has mostly done away with the norm-violating “tricks” of the past, in our contrived setting, victimless lying for candy is pervasive. And like the paired practice of costume-wearing and norm violations in past traditions, we see a relationship between costume and lying in our study. “Bad guys” in the no-salience condition lie more than “good guys”. This suggests that there is either a causal relationship or a self-selection effect of wearing a “bad guy” costume. Even if the effect is entirely through self-selection, this is still an interesting relationship. It means that those prone to norm violations select a consistent costumed identity.

Identity plays a second role in our study in the form of gender identity. We find that reporting a six increases the probability that a child later in line reports a six. This effect is only significant within gender. We offer a couple of reasons why this may be the case. First, if children tend to befriend within gender, trick-or-treat with friends, and emulate their friends, this could generate the result. Though we did not keep records, our impression was that most groups of trick-or-treaters were not groups of friends but families. This seems to be especially true for the younger trick-or-treaters, though there were certainly some groups of friends. However, the great majority of the trick-or-treaters within a participant’s group of ten would have been strangers even if they had approached the house with family or friends.

Indeed, many participated without any other individuals from an observable group, and those that participated with someone from their group usually only had no more than one or two accompanying them. Our sense is that the effect of friend emulation would have to be exceptionally strong to be the sole-driver of our within-gender peer effect. We offer another explanation for the within-gender peer effect: because socially acceptable behavior is often gender-specific, preferentially emulating others within one’s own gender is a simple and effective heuristic. Participants may have applied such a heuristic even when the norm does not vary by gender, as is the case for lying.

While we suggest caution in over-inferring from our boundary-case context (i.e. extreme costumes for children), the implications of our results suggest that policies that require particular kinds of dress may influence the behavior of the wearers. This has been a motivation for educators to adopt school uniforms and indeed Evans, Kremer and Ngatia (2008) found that school uniforms in Kenya reduced absenteeism. We speculate that the effect of clothing on an individual's sense of identity may also be a motivation for business dress codes. Wearing a suit serves to signal professionalism to clients, but in many businesses employees wear suits even on days for which they do not meet with clients. This suggests that such dress codes also serve an internal purpose. As a counterpoint, explicit "casual Friday" policies encourage less formal clothing as it is perceived to create a more convivial work environment.

We study the impact of Halloween costumes on the ethical behavior of 544 trick-or-treaters. We find that lying is more common when oneself is the beneficiary than when someone else is the beneficiary. We find that having the opportunity to lie for both oneself and someone else causes more honesty when reporting for oneself. Lying peaks at age 12 and is influenced by the lying of peers. In particular, peer effects are strong and predominantly within gender. Finally, we find that rendering the ethical orientation of one's costume salient leads to more or less lying depending on whether one's costume is a good guy or a bad guy. The salience of costume causes "good guys" to lie more and "bad guys" to lie less, consistent with a moral licensing. Though we believe this is the more plausible explanation, we cannot falsify the alternative explanation that trick-or-treaters believed that publicly reporting one's identity would cause differential monitoring from the experimenters.

This paper extends the literature on lying, and it is one of the first papers in economics that connects clothing and identity to ethical behavior. By leveraging a naturally-occurring cultural tradition — costume-wearing on Halloween — we elucidate the relationship between clothing and behavior using an extreme boundary case. The results also inform us on the nature of costume wearing itself, implying that the tradition may have served as a means to encourage norm-violations by temporarily changing people's sense of identity.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond, "Preferences for Truth-Telling," *Econometrica*, 2019, *87* (4), 1115–1153.

Adam, Hajo and Adam D. Galinsky, "Enclothed Cognition," *Journal of Experimental Social Psychology*, 2012, *48* (4), 918–925.

Akerlof, George A. and Rachel E. Kranton, "Identity and the Economics of Organizations," *Journal of Economic Perspectives*, March 2005, *19* (1), 9–32.

Ayal, Shahar, Jeremy Celse, and Guy Hochman, "Crafting messages to fight dishonesty: A field investigation of the effects of social norms and watching eye cues on fare evasion," *Organizational Behavior and Human Decision Processes*, 2019, *In Press.*

Belk, Russell W, "Halloween: An Evolving American Consumption Ritual," *ACR North American Advances*, 1990.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli, "Adaptive linear step-up procedures that control the false discovery rate," *Biometrika*, 2006, *93* (3), 491–507.

Bicchieri, Cristina, Eugen Dimant, and Silvia Sonderegger, "It's Not A Lie if You Believe the Norm Does Not Apply: Conditional Norm-Following with Strategic Beliefs," May 2020. Working Paper.

— , — , Simon Gächter, and Danielle Nosenzo, "Observability, Social Proximity, and the Erosion of Norm Compliance," March 2020. CESifo Working Paper, No. 8212, Center for Economic Studies and ifo Institute (CESifo), Munich.

Brocas, Isabelle and Juan D. Carrillo, "Self-serving, altruistic and spiteful lying in the schoolyard," *Working paper*, 2019.

Bucciol, Alessandro and Marco Piovesan, "Luck or cheating? A field experiment on honesty with children," *Journal of Economic Psychology*, 2011, *32*, 73–78.

Burke, Peter, *The Historical Anthropology of Early Modern Italy: Essays on Perception and Communication*, Cambridge University Press, 2005.

Cadsby, C Bram, Ninghua Du, and Fei Song, "In-group Favoritism and Moral Decision-making," *Journal of Economic Behavior & Organization*, 2016, *128*, 59–71.

Celse, Jérémy and Kirk Chang, "Politicians Lie, So Do I," *Psychological Research*, 2017, pp. 1–15.

Civile, Ciro and Sukhvinder Obhi, "Students Wearing Police Uniforms Exhibit Biased Attention towards Individuals Wearing Hoodies," *Frontiers in Psychology*, 2017, *8*, 62.

Clot, Sophie, Gilles Grolleau, and Lisette Ibanez, "Smug Alert! Exploring Self-licensing Behavior in a Cheating Game," *Economics Letters*, 2014, *123* (2), 191–194.

Cohn, Alain, Ernst Fehr, and Michel André Maréchal, "Business Culture and Dishonesty in the Banking Industry," *Nature*, 2014, *516* (729), 86–89.

_ , Michel André Maréchal, and Thomas Noll, "Bad Boys: How Criminal Identity Salience Affects Rule Violation," *The Review of Economic Studies*, 06 2015, *82* (4), 1289–1308.

DePaulo, Bella M, Deborah A Kashy, Susan E Kirkendol, Melissa M Wyer, and Jennifer A Epstein, "Lying in Everyday Life.," *Journal of Personality and Social Psychology*, 1996, *70* (5), 979.

Diekmann, Andreas, Wojtek Przepiorka, and Heiko Rauhut, "Lifting the Veil of Ignorance: An Experiment on the Contagiousness of Norm Violations," *Rationality and Society*, 2015, *27* (3), 309–333.

Dreber, Anna and Magnus Johannesson, "Gender Differences in Deception," *Economics Letters*, 2008, *99* (1), 197–199.

Erat, Sanjiv and Uri Gneezy, "White Lies," *Management Science*, 2012, *58* (4), 723–733.

Evans, David, Michael Kremer, and Muthoni Ngatia, "The Impact of Distributing School Uniforms on Children's Education in Kenya," Technical Report, World Bank March 2008.

Feil, DK, "How Venetians Think About Carnival and History," *The Australian Journal of Anthropology*, 1998, *9* (2), 141–162.

Feldman, Robert S, James A Forrest, and Benjamin R Happ, "Self-presentation and Verbal

Deception: Do Self-presenters Lie More?," *Basic and Applied Social Psychology*, 2002, *24* (2), 163–170.

Fischbacher, Urs and Franziska Föllmi-Heusi, "Lies in Disguise - An Experimental Study on Cheating," *Journal of the European Economic Association*, 2013, *11* (3), 525–547.

Gino, Francesca and Lamar Pierce, "Lying to Level the Playing Field: Why People May Dishonestly Help or Hurt Others to Create Equity," *Journal of Business Ethics*, 2010, *95* (1), 89–103.

_ , Shahar Ayal, and Dan Ariely, "Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel," *Psychological Science*, 2009, *20* (3), 393–398.

– , – , and _ , "Self-serving Altruism? The Lure of Unethical Actions that Benefit Others," *Journal of Economic Behavior & Organization*, 2013, *93*, 285–292.

Glätzle-Rützler, Daniela and Philipp Lergetporer, "Lying and age: An experimental study," *Journal of Economic Psychology*, 2015, *46*, 12–25.

Haley, Kevin J and Daniel MT Fessler, "Nobody's watching?: Subtle cues affect generosity in an anonymous economic game," *Evolution and Human behavior*, 2005, *26* (3), 245–256.

Jordan, Jennifer, Elizabeth Mullen, and J Keith Murnighan, "Striving for the Moral Self: The Effects of Recalling Past Moral Actions on Future Moral Behavior," *Personality and Social Psychology Bulletin*, 2011, *37* (5), 701–713.

Khan, Uzma and Ravi Dhar, "Licensing Effect in Consumer Choice," *Journal of Marketing Research*, 2006, *43* (2), 259–266.

Lasarov, Wassili and Stefan Hoffmann, "Social Moral Licensing," *Journal of Business Ethics*, 2018, *0*(0), 0.

Linton, Ralph and Ailin Linton, "Halloween through Twenty Centuries," 1950.

Maggian, Valeria and Marie Claire Villeval, "Social preferences and lying aversion in children," *Experimental Economics*, 2016, *19* (3), 663–685.

Mazar, Nina and Chen-Bo Zhong, "Do Green Products Make Us Better People?," *Psychological Science*, 2010, *21* (4), 494–498.

_ , On Amir, and Dan Ariely, "The Dishonesty of Honest People: A Theory of Self-concept Maintenance," *Journal of Marketing Research*, 2008, *45* (6), 633–644.

Michailidou, Georgia and Valentina Rotondi, "I'd Lie for You," *European Economic Review*, 2019, *118*, 181–192.

Miller, Kimberly A., Cynthia R. Jasper, and Donald R. Hill, "Costume and the Perception of Identity and Role," *Perceptual and Motor Skills*, 1991, *72* (3), 807–813.

Mol, Jantsje M, Eline C van der Heijden, and Jan JM Potters, "(Not) alone in the world: Cheating in the presence of a virtual observer," *Experimental Economics*, January 2020.

National Retail Federation, "Halloween Headquarters Survey: Halloween Spending to Reach $9 Billion," 2018. from https://nrf.com/insights/holiday-and-seasonal-trends/halloween.

Nieken, Petra and Simon Dato, "Compensation and Honesty: Gender Differences in Lying," 2016.

Nisan, Mordecai and Gaby Horenczyk, "Moral Balance: The Effect of Prior Behavior on Decision in Moral Conflict," *British Journal of Social Psychology*, 1990, pp. 29–42.

Pfattheicher, Stefan, Simon Schindler, and Laila Nockur, "On the impact of Honesty-Humility and a cue of being watched on cheating behavior," *Journal of Economic Psychology*, March 2019, *71*, 159–174.

Sachdeva, Sonya, Rumen Iliev, and Douglas L Medin, "Sinning Saints and Saintly Sinners: The Paradox of Moral Self-regulation," *Psychological Science*, 2009, *20* (4), 523–528.

Sanchez, Jafeth E., Andrew Yoxsimer, and George C. Hill, "Uniforms in the Middle School: Student Opinions, Discipline Data, and School Police Data," *Journal of School Violence*, 2012, *11* (4), 345–356.

Secilmis, Erdem, "An Experimental Analysis of Moral Self-regulation," *Applied Economics Letters*, 2018, *25* (12), 857–861.

Simonsohn, Uri, "Two Lines Test App," http://webstimate.org/twolines/ 2017.

—, "Two lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions," *Advances in Methods and Practices in Psychological Science*, 2018, *1* (4), 538–555.

Sterba, Richard, "On Hallowe'en," *American Imago*, 1948, *5* (3), 213–224.

Veblen, Thorstein, "The theory of the leisure class.(republished, 1934, by new york: Modern library)," 1899.

Wakin, Malham M, *Integrity First: Reflections of a Military Philosopher*, Lexington Books, 2000.

Walker, Jonathan, "Gambling and Venetian Noblemen c. 1500 - 1700," *Past & Present*, 02 1999, *162* (1), 28–69.

Ward, Donald J, "Halloween: An Ancient Feast of the Dead that Will Not Die," *Folklore and Mythology*, 1981, *1* (1), 4–6.

Wiltermuth, Scott S, "Cheating More When the Spoils are Split," *Organizational Behavior and Human Decision Processes*, 2011, *115* (2), 157–168.

Winkler, Louis and Carol Winkler, "Thousands of Years of Halloween," *New York Folklore Quarterly*, Sep 01 1970, *26* (3), 204.