Journal of Economic Behavior and Organization 179 (2020) 460–474

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Trick for a treat: The effect of costume, identity, and peers on norm violations☆

Shanshan Zhang, Matthew Gomies, Narek Bejanyan, Zhou Fang, Jason Justo, Li-Hsin Lin, Rainita Narender, Joshua Tasoff*

Department of Economic Sciences, Claremont Graduate University, 160 E 10th St, Claremont, CA 91711, USA

ARTICLE INFO

Article history:
Received 23 February 2020
Revised 12 July 2020
Accepted 11 September 2020

Keywords:
Lying
Identity
Costume
Clothes
Norm violation

ABSTRACT

We hypothesize that clothes can affect the behavior of the wearer by influencing the person's identity. We test this hypothesis by recruiting trick-or-treaters during Halloween, a time of year when people wear salient and extreme clothing. Because the tradition of costume-wear for Halloween evolved, in part, to hide one's identity during "tricks" (i.e. norm violations), we measure the effect of Halloween costumes on ethical behavior. We use the lying game of Fischbacher and Föllmi–Heusi as our experimental paradigm with $2 \times 3 \times 2$ conditions. First, we vary the stakes to price lying behavior. Second, we run three conditions with different beneficiaries of the report (self, other, and both) to test whether lying for others is perceived to be normative. Third, we manipulate the salience of one's costume to test the effect of costume and identity on ethical behavior. Surprisingly, we find that costume salience caused "good guys" to lie more and "bad guys" to lie less. We interpret this either as a moral licensing effect or as stemming from a perception of being monitored. Our design allows for the identification of contagion effects, and although there were no direct effects of gender, we found that children lie more when children of the same gender near them lie more. We also find that stakes had no effect, people lied more for themselves than for others, and lying has an inverted-U pattern over age, peaking at age 12.

1. 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 would like to thank Ibrahim Alamir, Bobae Hong, Kichang Kim, Uyen Le, Minh Pham, Yuxin Su, and Minjae Yun for helping us run the experiment.
* Corresponding author.
E-mail address: joshua.tasoff@cgu.edu (J. Tasoff).

https://doi.org/10.1016/j.jebo.2020.09.004
0167-2681/

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. In the next section we relate our study to the literature on lying, especially as a function of age, stakes, beneficiary and peers. We also discuss the literature on clothing, costumed festivals, identity and norm violations. The experiment design is described in Section 3. Section 4 contains the results. Sections 5 and 6 contain the Discussion and Conclusion, respectively.

2. 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 et al., 2019, for a meta-analysis). Notably, Cadsby et al. (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 childrens' 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 et al. (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 et al. (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 et al. (2015) and Bicchieri et al. (2020b) 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. (2020a) 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 et al. (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 et al. (2019) summarize the theory literature on Fischbacher and Föllmi-Heusi (2013) paradigm. They categorize models into those that have lying costs 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 do not observe or directly manipulate self-perceived reputation in our study, so this latter point is speculative. We did not 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.[1] 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 of identification occurs. Children often select characters that are aspirational or characters 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 et al. (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 et al. (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 et al. (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 et al. (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".

[1] We thank an anonymous referee for this important observation.

3. Research design

3.1. 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 1000 trick-or-treaters. Trick-or-treaters who approached the house between 6:00 and 9:30p.m. 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.

3.2. 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.

3.3. 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.

3.4. 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

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			

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.

3.5. 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 et al. (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.

4. Results

4.1. 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

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

no-salience and salience conditions with only a 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.

4.2. 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 second-order polynomial in order to fit an age trend. The coefficient on age is positive and significant while the coefficient on age^2 is negative, indicating that lying exhibits an inverted U pattern.[2] Fig. 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.

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

[2] Simonsohn (2018) suggests using the "two-line test" for testing whether two fitted lines have opposite signed slopes. The test confirms our inverted U finding. The results are available in Appendix A.

Table 3
Reporting a six – logistic regression with average marginal effects.

	(1)		(2)		(3)		(4)	
High Stakes	−0.005	*0.926*	−0.012	*0.809*	−0.002	*0.958*	−0.001	*0.981*
	(0.051)	[0.657]	(0.05)	[0.657]	(0.04)	[0.657]	(0.04)	[0.657]
Other	−0.131**	*0.025*	−0.137**	*0.018*	−0.115**	*0.015*	−0.117**	*0.013*
	(0.058)	[0.087]	(0.058)	[0.078]	(0.047)	[0.074]	(0.047)	[0.074]
Both	−0.13*	*0.062*	−0.14**	*0.046*	−0.092	*0.109*	−0.094	*0.107*
	(0.07)	[0.111]	(0.07)	[0.093]	(0.058)	[0.125]	(0.058)	[0.125]
Other × Both	0.012	*0.885*	0.018	*0.828*	−0.016	*0.832*	−0.017	*0.827*
	(0.082)	[0.657]	(0.082)	[0.657]	(0.076)	[0.657]	(0.076)	[0.657]
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 × Female Peer Proportion							0.036	*0.865*
							(0.213)	[0.657]
Male Peer Proportion							0.285***	*0.004*
							(0.1)	[0.046]
Female × 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]
N		685		685		685		685
Clusters		55		55		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. 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

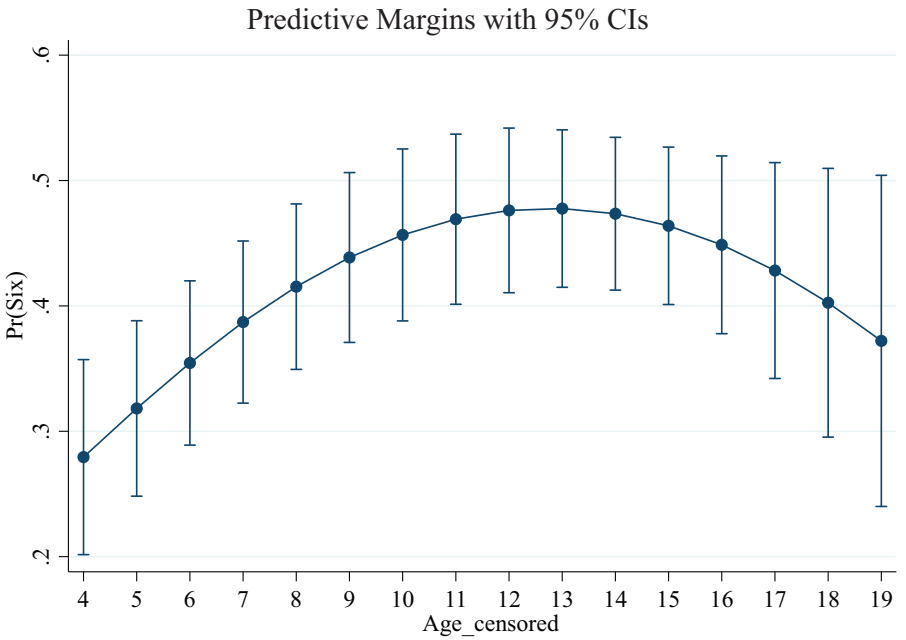

Fig. 1. Reporting a six as a function of age. *Notes:* Probability of reporting a six across ages. Error bars indicate 95% confidence intervals. Ages are censored at 4 and 19.

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 × 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 × Bad Guy	−0.169**	*0.036*	−0.268***	*0.003*
	(0.08)	[0.093]	(0.089)	[0.039]
Salience × Other			−0.168**	*0.044*
			(0.083)	[0.093]
Bad Guy × Other			−0.239**	*0.03*
			(0.11)	[0.092]
Salience × Bad Guy × 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

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 Propotion" 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 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

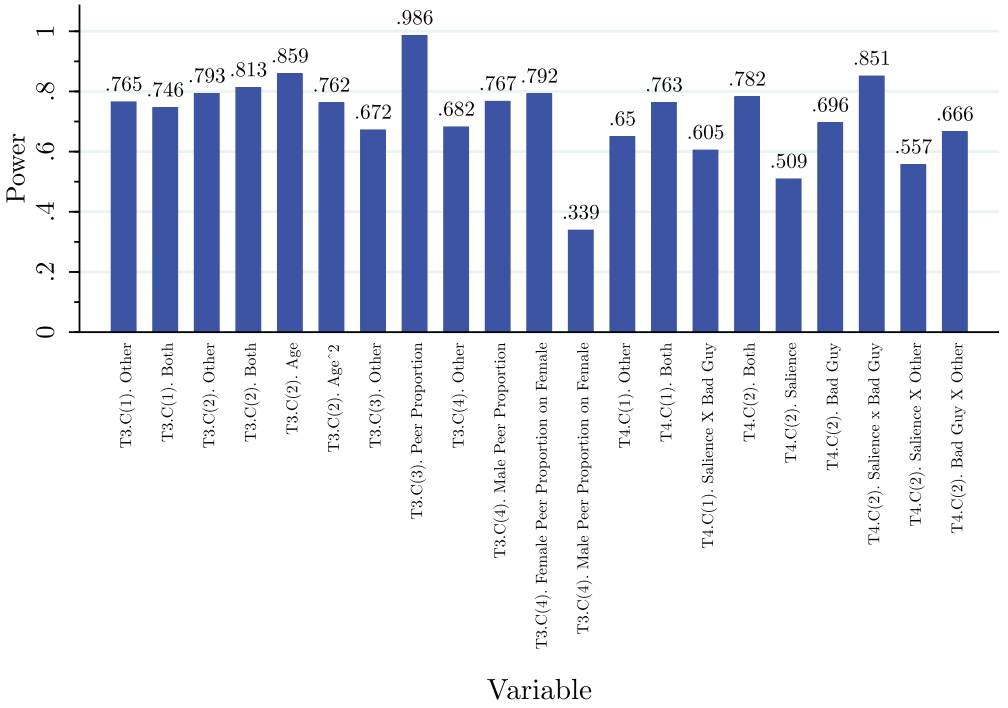

Fig. 2. Reporting a six as a function of age. *Notes:* "Tx.C(*y*). variable" refers to the coefficient of the variable found in Table *x*, Column (*y*). The values displayed in the figure show the statistical power of finding an effect of the size we estimated with critical value of $\alpha = 0.05$. This figure includes all tests for which the value was significant at $p < 0.1$. See main text for how the values are simulated.

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.

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.

4.3. 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,[3] 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 1000 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).

Fig. 2 displays our estimated statistical power for each coefficient of interest that has $p < 0.1$ in the paper. Statistical power for most of our main results exceeds 0.7. The other condition reduces lying and we find that the power ranges from

[3] For example, comparing beneficiaries tests between three conditions: self, other, and both. Since we pool the other orthogonal treatments within these three group, there are approximately 180 in each of these three conditions.

0.65 to 0.793. The both condition ranges in power from 0.746 to 0.859. Our age results are presented in Col (2) and have a power of 0.859 and 0.762 on the linear and quadratic terms respectively. Turning to the peer effect in Col (3), we find that the test is powered at 0.986. When we decompose the peer effects by gender we find that the male peer effect on males is powered at 0.767 and the female effect on females is powered at 0.792.

In Table 4, our main specification is Col (2) and the two main variables of interest are Salience and Salience × Bad Guy, which are positive and negative respectively. These two coefficients are powered at 0.59 and 0.851 respectively. While there is no conventional threshold for statistical power in the same way there is for critical values, casual conversations amongst experimentalists suggest that a power of 0.8 is considered good.

5. Discussion

We found that our high-stakes manipulation had no effect. This is in-line with the results of previous studies (Mazar et al., 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 et al. (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 et al. (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 et al., 2014; 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 et al. (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 et al., 2019; Ayal et al., 2020), 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.[4]

[4] We hoped to conduct a follow-up experiment in which trick-or-treaters would be primed with the same question, "Is (blank) a good guy or a bad guy?", but would not need to report it aloud. Unfortunately, the global pandemic of 2020 interfered with the ability to conduct follow-up Halloween experiments.

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 et al. (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.

Discussion points: 1. Why did we get moral licensing instead of behavior in congruity with clothes? (a) Maybe they did congruent things before. (b) Maybe because it was a public announcement of identity to a watching adult. 2. Age, does candy matter as much? 3. Interesting that the selection effect goes in opposite direction of priming effect.

One pivotal finding from this study is the incongruent nature of lying outcomes under the priming condition. Our a priori hypothesis proposed trick-or-treaters will be invested in their costume choice and subsequently a reinfornced priming of such costume will induce actions congruent with their costume's theme:good guy or bad guy. More contextually, we expected individuals dressed as "good guys" would lie less when given the opportunity to lie for greater returns, candies. In actuality, our results indicate that primed trick-or-treaters dressed as "bad guys" lied less for candy compared to their "good guy" counterparts. [This is not correct]. A proposed explanation to this observed phenomena is moral licensing: "good guys" feel their costume equates to moral tokens providing them less incentive to follow norms and more incentive to lie making use of their good deed tokens. Conversely, given that "bad guys" have now been observed as category "bad" by the experimenter conducting the primer, they may feel persuaded to lie less in order to reinforce and overtly show their true identity does not align with their "bad guy" costume. When comparing priming vs. non-priming groups, the data indicates that "good guys" lying for more candies is more frequent for primed conditions versus non-primed conditions.

One of the intriguing finding that we get is the counter-expecting priming effect. Intuitively, after we prime the participants to focus on their costume, they would be more invest in their character role playing – expecting lying less for good guys and lying more for bad guys. This is also the mechanism behind in custom, identity, and norm violation behavior in our theoretical context. However, the results turn out to be opposite. This does not mean the link has been torn down, but maybe just two sides of the same story. The rationale we have here is moral licensing: the good guys lie more since they get moral licensed. Yet still there might exist some "moderator" that we overlooked. One of the differences between laboratory experiment and field experiment is "context". The context here is that we take advantage of Halloween costume and trick or treat tradition to delve in act in identity and norm violation, which a little mischievous behavior is infact following alone with social expectation. Maybe integrity good guy concept obsolete (Yes, Iron man rock compares to the Captain); Maybe good guys' player suppose they should get bonus candy for the effort of saving world. This is the part when moral licensing jumps in, because good guys lie less for themselves without priming but it increases for the group that primed. Nevertheless, moral licensing says little about bad guys behavior. We believe some characters to those kids may not be all bad, which means when chances come, they might want to clear that stigma for them. Cosplay is about role-playing and honoring based on passion to the character. There is possibility that through priming, we trigger the acting part of them to extend the story of their character. The most far-reaching discussion would related to why these children chose these costume and if this related to their understanding to social expectation, social norm, and norm violation behavior, but for here, we attribute this beautiful cross-over interaction to Halloween trick or treat context.

[Several sample selection and experimental issues are important to note. First, there might exist selection bias since the costumes were pre-selected by the trick-or-treaters. Wearing a "bad guy" costume increases the probability of reporting a six for one's self. It is possible that wearing a "bad guy" costume affects wearer's lying behavior. But It is also possible that

a person who is more likely to lie is also more likely to chose a "bad guy" costume. However, this does not pose a problem for inferring the effect of priming since priming was randomized in our design. It would be interesting to see whether the result will be different, if the clothes are given to them, instead of wearing the pre-selected costumes.

Naturally, as with any empirical exercise, questions of external validity remain. In particular, whether results of this experiment were particular to the context of trick-or-treater during Halloween, whether candy provides sufficient incentive. For instance, we believe how many candies you get during Halloween night matters to a 8 years old kid, but how much it matters to a 15 years old teenager. It is possible that lying declines after age of 12 or 13 due to less motivation of getting more candies.

Another concern is that the moral licensing effect of good guys might be caused by the prior good deeds that they did while wearing the "good guys" costume before coming to the trick-or-treating event on the Halloween night. More research is needed to find the reason why we found moral licensing effect instead of behavior in congruity with clothes.

Interestingly, in the priming condition, "bad guys" lied less compared to the un-primed "bad guys". This result might be caused by the public announcement of "bad guys" identity. When subjects reported "bad guys" to us, they might be afraid to lie since "bad guys" have the tendency to be watched or judged. It might be a different result if we priming them without reporting "good or bad" to us.

Last, due to the complexity of the design where the kids need to roll two dice and reported twice for both treatment, the decrease in the probability of lying can be caused by the heavy cognitive loading for the kids. Thus, it would also be interesting to study whether their lying behavior changes if they roll only one die that determines the payoff for both themselves and others, instead of one die for each.

6. Conclusion

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.

Author's contributions

SZ led the project. MG contributed a large fraction of the analysis. JT is the principal investigator. All other authors contributed equally to the project. There are no conflicts of interest from any of the authors.

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.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Simonsohn (2018) proposed the two lines test to detect U-shapes without a functional form assumption. The basic idea is to estimate two regression lines, one for low values and one for high values of the independent variable, and then verify that the two slopes have an opposite sign and are individually statistically significant. The breakpoint between the lines is discovered algorithmically.

We use Simonsohn's app (Simonsohn, 2017), which finds an inverted U-shaped effect of age variable to the probability of rolling a six, while controlling other variables such as the stakes, beneficiaries, female, and parents. The average slope of the effect of age below the age of 12 is 0.03 (p=0.0021) and the average slope of the effect of age above the age of 12 is −0.03 (p=0.012). Please refer to Fig. A.1 for more details.

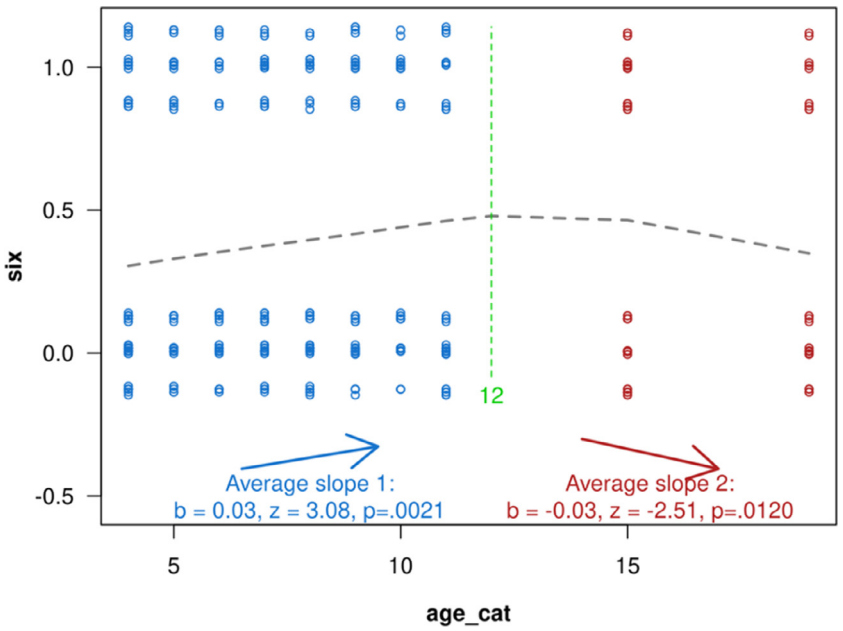

Fig. A.1. Two-lines Test for Inverted U-shaped Age Effects. *Notes:* Probability of reporting a six across ages. Ages are censored at 4 and 19.

References

Abeler, J., Nosenzo, D., Raymond, C., 2019. Preferences for truth-telling. Econometrica 87 (4), 1115–1153. doi:10.3982/ecta14673.
Adam, H., Galinsky, A.D., 2012. Enclothed cognition. J. Exp. Soc. Psychol. 48 (4), 918–925. doi:10.1016/j.jesp.2012.02.008.
Akerlof, G.A., Kranton, R.E., 2005. Identity and the economics of organizations. J. Econ. Perspect. 19 (1), 9–32. doi:10.1257/0895330053147930.
Ayal, S., Celse, J., Hochman, G., 2019. Crafting messages to fight dishonesty: a field investigation of the effects of social norms and watching eye cues on fare evasion. Org. Behav. Hum. Decis. Process. (In Press).
Belk, R.W., 1990. Halloween: an evolving american consumption ritual. ACR North Am. Adv..
Benjamini, Y., Krieger, A.M., Yekutieli, D., 2006. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93 (3), 491–507.
Bicchieri, C., Dimant, E., Gächter, S., Nosenzo, D., 2020a. Observability, social proximity, and the erosion of norm compliance. CESifo Working Paper, No. 8212, Center for Economic Studies and ifo Institute (CESifo), Munich.
Bicchieri, C., Dimant, E., Sonderegger, S., 2020b. It's not a lie if you believe the norm does not apply: conditional norm-following with strategic beliefs. Working Paper.
Brocas, I., Carrillo, J. D., 2019. Self-serving, altruistic and spiteful lying in the schoolyard. Working Paper.
Bucciol, A., Piovesan, M., 2011. Luck or cheating? A field experiment on honesty with children. J. Econ. Psychol. 32, 73–78.
Burke, P., 2005. The Historical Anthropology of Early Modern Italy: Essays on Perception and Communication. Cambridge University Press.
Cadsby, C.B., Du, N., Song, F., 2016. In-group favoritism and moral decision-making. J. Econ. Behav. Org. 128, 59–71.
Celse, J., Chang, K., 2017. Politicians lie, so do I. Psychol. Res. 1–15.
Civile, C., Obhi, S., 2017. Students wearing police uniforms exhibit biased attention towards individuals wearing hoodies. Front. Psychol. 8, 62.
Clot, S., Grolleau, G., Ibanez, L., 2014. Smug alert! exploring self-licensing behavior in a cheating game. Econ. Lett. 123 (2), 191–194.
Cohn, A., Fehr, E., Maréchal, M.A., 2014. Business culture and dishonesty in the banking industry. Nature 516 (729), 86–89. doi:10.1038/nature13977.
Cohn, A., Maréchal, M.A., Noll, T., 2015. Bad boys: how criminal identity salience affects rule violation. Rev. Econ. Stud. 82 (4), 1289–1308. doi:10.1093/restud/rdv025.
DePaulo, B.M., Kashy, D.A., Kirkendol, S.E., Wyer, M.M., Epstein, J.A., 1996. Lying in everyday life.. J. Person. Soc. Psychol. 70 (5), 979.
Diekmann, A., Przepiorka, W., Rauhut, H., 2015. Lifting the veil of ignorance: an experiment on the contagiousness of norm violations. Rational. Soc. 27 (3), 309–333. doi:10.1177/1043463115593109.
Dreber, A., Johannesson, M., 2008. Gender differences in deception. Econ. Lett. 99 (1), 197–199.
Erat, S., Gneezy, U., 2012. White lies. Manag. Sci. 58 (4), 723–733.
Evans, D., Kremer, M., Ngatia, M., 2008. The impact of distributing school uniforms on children's education in Kenya. Technical Report. World Bank.
Feil, D., 1998. How venetians think about carnival and history. Austral. J. Anthropol. 9 (2), 141–162.
Feldman, R.S., Forrest, J.A., Happ, B.R., 2002. Self-presentation and verbal deception: do self-presenters lie more? Basic Appl. Soc. Psychol. 24 (2), 163–170.
Fischbacher, U., Föllmi-Heusi, F., 2013. Lies in disguise – an experimental study on cheating. J. Eur. Econ. Assoc. 11 (3), 525–547. doi:10.1111/jeea.12014.
Gino, F., Ayal, S., Ariely, D., 2009. Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. Psychol. Sci. 20 (3), 393–398. doi:10.1111/j.1467-9280.2009.02306.x.
Gino, F., Ayal, S., Ariely, D., 2013. Self-serving altruism? The lure of unethical actions that benefit others. J. Econ. Behav. Org. 93, 285–292.
Gino, F., Pierce, L., 2010. Lying to level the playing field: why people may dishonestly help or hurt others to create equity. J. Bus. Ethics 95 (1), 89–103.
Glätzle-Rützler, D., Lergetporer, P., 2015. Lying and age: an experimental study. J. Econ. Psychol. 46, 12–25. doi:10.1016/j.joep.2014.11.002.
Haley, K.J., Fessler, D.M., 2005. Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. Evol. Hum. Behav. 26 (3), 245–256.
Jordan, J., Mullen, E., Murnighan, J.K., 2011. Striving for the moral self: the effects of recalling past moral actions on future moral behavior. Personal. Soc. Psychol. Bull. 37 (5), 701–713.
Khan, U., Dhar, R., 2006. Licensing effect in consumer choice. J. Mark. Res. 43 (2), 259–266.
Lasarov, W., Hoffmann, S., 2018. Social moral licensing. J. Bus. Ethics 0 (0), 0. doi:10.1007/s10551-018-4083-z.
Linton, R., Linton, A., 1950. Halloween through twenty centuries.
Maggian, V., Villeval, M.C., 2016. Social preferences and lying aversion in children. Exp. Econ. 19 (3), 663–685. doi:10.1007/s10683-015-9459-7.
Mazar, N., Amir, O., Ariely, D., 2008. The dishonesty of honest people: a theory of self-concept maintenance. J. Mark. Res. 45 (6), 633–644.

Mazar, N., Zhong, C.-B., 2010. Do green products make us better people? Psychol. Sci. 21 (4), 494–498.
Michailidou, G., Rotondi, V., 2019. I'd lie for you. Eur. Econ. Rev. 118, 181–192. doi:10.1016/j.euroecorev.2019.05.014.
Miller, K.A., Jasper, C.R., Hill, D.R., 1991. Costume and the perception of identity and role. Percep. Motor Skills 72 (3), 807–813. doi:10.2466/pms.1991.72.3.807.
Mol, J.M., van der Heijden, E.C., Potters, J.J., 2020. (not) alone in the world: cheating in the presence of a virtual observer. Exp. Econ..
National Retail Federation, 2018. Halloween headquarters survey: halloween spending to reach $9 billion. From https://nrf.com/insights/holiday-and-seasonal-trends/halloween.
Nieken, P., Dato, S., 2016. Compensation and honesty: gender differences in lying.
Nisan, M., Horenczyk, G., 1990. Moral balance: the effect of prior behavior on decision in moral conflict. Br. J. Soc. Psychol. 29–42.
Pfattheicher, S., Schindler, S., Nockur, L., 2019. On the impact of honesty-humility and a cue of being watched on cheating behavior. J. Econ. Psychol. 71, 159–174.
Sachdeva, S., Iliev, R., Medin, D.L., 2009. Sinning saints and saintly sinners: the paradox of moral self-regulation. Psychol. Sci. 20 (4), 523–528.
Sanchez, J.E., Yoxsimer, A., Hill, G.C., 2012. Uniforms in the middle school: student opinions, discipline data, and school police data. J. School Viol. 11 (4), 345–356. doi:10.1080/15388220.2012.706873.
Secilmis, E., 2018. An experimental analysis of moral self-regulation. Appl. Econ. Lett. 25 (12), 857–861. doi:10.1080/13504851.2017.1374530.
Simonsohn, U., 2017. Two lines test app. http://webstimate.org/twolines/.
Simonsohn, U., 2018. Two lines: a valid alternative to the invalid testing of u-shaped relationships with quadratic regressions. Adv. Methods Pract. Psychol. Sci. 1 (4), 538–555.
Sterba, R., 1948. On hallowe'en. Am. Imago 5 (3), 213–224.
Veblen, T., 1899. The theory of the leisure class.(republished, 1934, by new york: Modern library).
Wakin, M.M., 2000. Integrity First: Reflections of a Military Philosopher. Lexington Books.
Walker, J., 1999. Gambling and Venetian Noblemen c. 1500 – 1700. Past Present 162 (1), 28–69. doi:10.1093/past/162.1.28.
Ward, D.J., 1981. Halloween: an ancient feast of the dead that will not die. Folklore Mythol. 1 (1), 4–6.
Wiltermuth, S.S., 2011. Cheating more when the spoils are split. Org. Behav. Hum. Decis. Process. 115 (2), 157–168.
Winkler, L., Winkler, C., 1970. Thousands of years of halloween. New York Folklore Q. 26 (3), 204.