NUFFIELD CENTRE FOR EXPERIMENTAL SOCIAL SCIENCES

DISCUSSION PAPER SERIES

2011 - 004

Lifting the veil of ignorance: An experiment on the contagiousness of norm violations

Andreas Diekmann (ETH Zurich)

Wojtek Przepiorka (University of Oxford)

Heiko Rauhut (ETH Zurich)

July 2011

Lifting the veil of ignorance: An experiment on the contagiousness of norm violations

Andreas Diekmann[a,1], Wojtek Przepiorka[a,b,1], Heiko Rauhut[c,1,2]

[a] ETH Zurich, Chair of Sociology, Clausiusstrasse 50, CLU D 4, CH-8092 Zurich, Switzerland

[b] University of Oxford, Department of Sociology, Manor Road, Oxford OX1 3UQ, UK

[c] ETH Zurich, Chair of Sociology, in particular of Modeling and Simulation, Clausiusstrasse 50, CLU E 6, CH-8092 Zurich, Switzerland

[1] All authors contributed equally to this work.

[2] To whom correspondence may be addressed: rauhut@gess.ethz.ch

July 1, 2011

Abstract

Norm violations can be contagious. Previous research analyzed two mechanisms of why knowledge about others' norm violations triggers its spread: (1) Actors lower their subjective beliefs about the probability or severity of punishment, or (2) they condition their compliance on others' compliance. While earlier field studies could hardly disentangle both effects, we use a laboratory experiment which eliminated any punishment threat. Subjects (n = 466) could commit a violation of the honesty norm. They threw a die and were paid according to their reported number. Our design ruled out any possibility of personal identification so that subjects could lie about their thrown number and claim inflated payoffs without risking detection. The aggregate distribution of reported payoffs allowed determining the extent of liars in the population. Two treatments in which subjects were informed about lying behavior of others were compared to a control condition without information feedback. Distributions from a subsequent dice throw revealed that knowledge about liars triggered the spread of lying compared to the control condition. Our results demonstrate the contagiousness of norm violations, where actors imitate norm violations of others under the exclusion of strategic motives.

Keywords: Social norms; social influence; cheating; lying; experimental sociology; conditional cooperation; Heinrich Popitz

I. Introduction

Ignorance can act as a protective barrier to the spread of norm violations. This hypothesis has its German roots in the writings of Heinrich Popitz (1968). In the American sociological tradition, a similar notion has been raised by the so-called broken windows theory. Both traditions emphasize the contagiousness of norm violations. The paradigmatic example is the observation of broken windows in a neighborhood or other signs of disorder, which eventually trigger the spread of norm violations (Wilson 1982).

There is a growing body of empirical research testing the hypothesis, an example of which is the series of field experiments by Keizer et al. (2008). These experiments tested whether information about a certain kind of disorder triggered the spread of other kinds of disorder. The authors could show that signs of graffiti caused people to litter, public occurrences of illegal parking caused illegal trespassing, a large extent of unreturned shopping carts at a supermarket caused littering and signs of graffiti and litter caused people to steal money from sealed envelopes hanging out of mailboxes. In another set of field experiments, Cialdini et al. (1990) showed that people have a higher propensity to litter into the environment if they observe others littering and if the environment is already littered.

While the empirical evidence for the contagiousness of norm violations is fairly robust, the causal mechanisms driving this effect are unclear. Two main factors can be identified, which are rooted in two different schools of thought. First, the effect could be explained by *rational-choice theory*. Observing others' norm violations may cause people to update their beliefs regarding their cost-benefit analysis of norm violations (Becker 1968, Allingham and Sandmo 1972, Bikhchandani, Hirshleifer and Welch 1998, Groeber and Rauhut 2010). Signs of disorder could therefore show that norm violations are hardly detected and, if so, only mildly punished.

Second, the effect could be explained by the social psychological focus theory (Cialdini et al. 1990; see also Gino et al. 2009). Observations of others' norm violations may change actors' beliefs about the appropriateness of their own actions. In this sense, descriptive norms, i.e. what most people do, affect injunctive norms, i.e. what most people approve of. To put it in simple words: if others break a certain norm, it is fine if you break it too. We call this effect *conditional norm compliance.* More precisely, we mean with conditional norm compliance the motivation of individuals to adhere to norms if they believe that others do so as well and to violate them if they believe that others violate them.[1]

[1] The concept of conditional norm compliance relates to the recent debate about *conditional cooperation* in experimental economics (Fischbacher et al. 2001; Gächter 2007). However, the concept is more general in so far as it refers to pure imitation without requiring strategic motives of reciprocity.

This paper seeks to disentangle both mechanisms by ruling out one of them by design. Our objective is to demonstrate that conditional norm compliance is sufficient to trigger the effects of ignorance on norm compliance. While other studies, like the recent one by Keizer, Lindenberg and Steg (2008) or the ones in the broken windows tradition (Sampson and Raudenbush 2004; Wilson and Kelling 1982) had to come up with additional assumptions that punishment of observed norm violations is simply uncommon and therefore negiglible, we can rule out the punishment argument not only by arm-chair speculation but by our experimental design. In what follows, we discuss examples, case studies and previous research regarding the contagiousness of norm violations. Then we introduce our experimental design, analyze our empirical results and discuss our findings.

II. CONTAGIOUSNESS OF NORM VIOLATIONS

Imagine you knew that most inhabitants of your city evaded taxes, cheated on their partners, dodged paying the fares for public transport, lied about their age at dates and crossed against red lights. While we may intuitively think that a high detection rate of norm violations maintained social order, this very knowledge of others' norm violations may lower the willingness to comply with norms. Loosely based on the aphorism "what the eye does not see, the heart does not grieve over", ignorance of norm violations can have a preventive effect. In what follows, we use the term *ignorance hypothesis* to refer to the preventive effect of not knowing about others norm violations.

We know the contagious effects of norm violations from scenarios in which norm violations are visible. If many people start to cross against the red light at a crosswalk, others join in. If many cars in a street park in clearways, the sideway and other no-parking zones quickly become jammed of cars. Seeing a littered beach makes people more likely to litter themselves, which gradually pollutes the beach. However, many norm violations are not directly visible but conducted in private. Two-timing, tax evasion, consumption of pornography, visits of prostitutes or alcohol abuse are behaviors which are often concealed from others. Some of which are revealed, but others remain covered.

Seneca was probably one of the first making the observation that ignorance can act as a protective barrier to the spread of norm violations. About two thousand years ago, he wrote in his Moral Essays:

> "In that state in which men are rarely punished a sympathy for uprightness is formed, and encouragement is given to this virtue as to a common good. Let a state think itself blameless, and it will be so; its anger against those who depart from the general sobriety will be greater if it sees that they are few. Believe me, it is dangerous to show a state in how great a majority evil men are." (Seneca 1928 [first 63], xxii. 3-xxiv. 1)[2]

Nearly 2000 years later, the German sociologist Popitz (1968) suggested a more detailed account specifying the *ignorance hypotheses* by three impossibility statements:

> "Impossible is the complete transparency of all norm-relevant behaviors in society, a normative system which could cope with the detection of all norm violations, and a punishment system which would retain its protective function if all norm violations were sanctioned" (Popitz 1968, 18, translated by the authors).

Popitz (1968) illustrates the ignorance hypotheses by a thought experiment from Thackeray's (1869) classical novel. Consider what would happen if every norm violation was actually detected:

> "Just picture to yourself everybody who does wrong being found out, and punished accordingly. Fancy all the boys in all the schools being whipped; and then the assistants, and then the headmaster (…) Fancy the provost marshal being tied up, having previously superintended the correction of the whole army. (…) The butchery is too horrible. The hand drops powerless, appalled at the quantity of birch which it must cut and brandish. I am glad we are not all found out" (Thackeray, 1869, as quoted in Popitz 1968).

The impossibility of complete transparency of normative behavior builds the foundation for two macro-sociological hypotheses: 1. If the complete extent of norm violations was known, norm violations would gradually increase and the normative system would collapse. 2. If all norm violations were sanctioned, the punishment system would collapse. The counterintuitive

[2] These observations built the basis of Seneca's political recommendation to uphold ignorance of the number of slaves: "A proposal was once made in the senate to distinguish slaves from free men by their dress; it then became apparent how great would be the impending danger if our slaves should begin to count our number. Be sure that we have a like danger to fear if no man's guilt is pardoned; it will soon become apparent how greatly the worse element of the state preponderates." (Seneca 1928 [first 63], xxii. 3-xxiv. 1) More recently, in The New York Times, Friedman (2011) pointed out that, among other things, the information about the unequal distribution of land accessible through Google Earth lead to popular uprising against the ruling family in Bahrain. Note however that the question of how perceived inequality may deteriorate the legitimacy of an autocratic system is a related but different question than the one implied in our *ignorance hypothesis*. Here we suggest that ignorance about norm violations prevents the diffusion of such behavior and thus stabilizes a normative system.

point here is that the political goal to uncover undetected norm violations actually counteracts the stability of the norm.

It is possible to reconstruct the macro-sociological correlation of the ignorance hypothesis by micro-level mechanisms. Its main premise is that actors do not exactly know if and to what extent other members of society violate a particular norm – be it black labor, corruption, smuggling, shoplifting, "deviant" sexual practices or adultery: "There is in general a large discrepancy between the actual and the subjectively perceived extent of norm violations." (Popitz 1968: 15) Ignorance typically implies underestimation of the extent of norm violations in the population, which lowers actors' propensity for norm violations. "Many social forces work towards generating a relatively favorable representation of norm compliance in society, which may also be described as the general requirement for the 'functioning' of a particular social norm in society" (Popitz 1968:15). This implies that if the actual extent of norm violations was known, the legitimacy of the norm would deteriorate.

The aforementioned theoretical reasoning may be illustrated by an example from the history of the German Democratic Republic (GDR). The socialist state discouraged to watch Western television channels and political campaigns against households lacking this kind of compliance were launched. The crux was that households consuming Western television were identifiable by the direction of their antennas upon their roofs. Walter Ulbricht, the leader of the East Germany Communist Party at that time realized that this lifted the veil of ignorance and stated "the class enemy is sitting on the roof" (Spiegel 1980, p. 42). He tried to revert the situation by encouraging the East German youth organization "Freie Deutsche Jugend" to tear down television antennas oriented to receive Western German channels (Marks 1983, p. 50). Too late – the otherwise unknown extent of Western television consumption was already revealed to everybody. In fact, this knowledge eventually triggered the collapse of the prohibitive norm against Western television consumption in 1973 (Spiegel 1980, p. 44).

Furthermore, lifting the veil of ignorance may even trigger normative change as the Kinsey report (1948; 1953) about sexual behaviors in the United States exemplified:

> "In sum, Kinsey was the major factor in changing attitudes about sex in the twentieth century. (...) He changed the nature of sexual studies, forced a reexamination of public attitudes toward sex, challenged the medical and psychiatric establishment to reassess its own views, influenced both the feminist movement and the gay and lesbian movement,

> and built a library and an institution devoted to sex research. His reputation continues to grow, and he has become one of the legends of the twentieth century.“ (Bullough 1998, p. 131)

With respect to our argument, the veil of ignorance of various sexual practices was lifted. The extent of homosexual behavior, masturbation, oral sex and other practices became public. It became apparent that these behaviors were more widespread than previously thought, which triggered their subsequent spread and eventually contributed to a fundamental change in norms and values in people’s sex lives. We test the ignorance hypothesis in a simple laboratory experiment, the design of which will be explained next.

III. DESIGN OF THE EXPERIMENT

The question as to whether ignorance has a preventive effect is an empirical question. One problem with traditional research designs such as surveys or official crime records, however, is that the actual extent of norm violations is not only unknown to the population but also to the researcher. Generally, there is undercoverage of norm violations so that the actual rate has to be estimated by making a number of influential assumptions which are often hard to validate. In contrast, laboratory experiments enable to measure the complete rate of norm violations in a very direct way.

In laboratory experiments, however, it may be that the true rate of norm compliance is overestimated because subjects feel observed and may react overly norm adherent. These considerations led us to the implementation of a laboratory experiment which eliminated the possibility to identify which particular subject committed a norm violation. Our desired implementation would guarantee perfect anonymity on the individual level and allow to correctly inferring the extent of norm violations from aggregated data.

We chose the dice experiment of Fischbacher and Heusi (2008) as basic design of our study. Later on, we realized that the complete absence of a sanctioning threat was a desirable side effect of the design. The experiment enabled subjects to commit a violation of the honesty norm. Subjects threw a die in a completely anonymous setting, meaning that they were alone and not observed at all. Then, they had to report their number in complete privacy by entering it in a computer system located in an isolated box. Subjects received cash in Swiss Francs according to the number they reported. The only exception was the number six, which corresponded with zero earnings. All subjects knew that they were the only ones knowing the true number they actually threw. Therefore, subjects could distort the truth and report a number which yielded them higher monetary earnings. This meant that all subjects who did not throw a “five” faced an internal conflict between maximizing their own payoff by reporting a higher number and adhering to the honesty norm. The implementation of anonymity among subjects and between subjects and experimenter eliminated any possibility to detect liars. This guaranteed that our experimental design avoided any confounding with punishment threats, neither material nor social ones. In addition, our experiment avoided any confounding with strategic reasoning. Individual lying did not affect the payoffs of other subjects.

Of course, our experiment did not allow for the observation of individual norm violations. It was actually the point of the whole procedure to guarantee individuals’ complete anonymity for a valid elicitation of behaviors. Nonetheless, we can estimate the extent of norm violations at the group level. If the honesty norm was consistently followed by all subjects, we would observe a rate of about one sixth for every reported number. Therefore, the extent of norm violations can be measured by computing the difference between theoretical and observed rates for each number.[3]

According to the ignorance hypothesis, knowledge about norm violations in the population should trigger subsequently more norm violations. Therefore, we informed subjects about the distribution of reported numbers after the first throw. Thus, we uncovered the dark field of norm violations. We implemented two ways of communicating the empirical distributions of first throws. In the information condition one (info 1), we presented the distribution of the dice experiment of Fischbacher and Heusi (2008). Thus, we explained the subjects that the distribution they saw was an empirical distribution from 389 subjects who came from the same universities and previously participated in the same experiment (the distribution and the respective experimental instructions can be seen in Figures A4 and A5 in the appendix). The distribution has a clear pattern. The over-reporting of the numbers “four” and “five” and the implied violation of the honesty norm is apparent. In the information condition two (info 2), we reported the distribution of the group which was currently playing the game (see Figure A6).

[3] In this sense, the method is comparable to the so-called randomized response method (Warner 1965).

The advantage of the first information condition is that the distribution represents behaviors of over three hundred subjects. Further, it does not vary over different experimental sessions. The disadvantage is that subjects may consider the distribution as "external", because it stems from a group of different subjects. This is different for the "internal" implementation of information condition two, which represented the behaviors of the very subjects who participated in the same experimental session. Here, however, the dice throws naturally generated a great deal of randomness. This means that the generated group distribution in each session was always a different distribution generated from a different group of subjects. Because both methods are complementary in their advantages and disadvantages, they balance each other and enable robustness checks of our empirical results.

Further, we implemented a control treatment in which we did not present the distribution of first throws. After the first throws and the presentation of the respective distribution (respectively its absence in the control group), we implemented another round in which subjects were asked to throw the die again. We paid subjects for both throws similarly. Hence, the maximum payment was ten Swiss Francs. Table 1 summarizes our design.

Table 1: The experimental design

control group (info 0)	O		O
experimental group 1 (info 1)	O	X_1	O
experimental group 2 (info 2)	O	X_2	O

O denotes observation, X intervention.

The instructions made it clear that subjects were allowed to test their die. Thus, they were allowed to throw the die as many times as they wished. However, only the first throw counted. This rule was stated very explicitly. This setup made it possible for subjects to ensure that they were not deceived by fraud dice. Furthermore, this setup implicitly provided self-justifications for lying and therefore made it easier for subjects to violate the honesty norm in favor of their self-interest. Subjects who were unlucky with their first throw might continue to throw the die and be tempted to report another, more favorable number which occurred later on. Potentially, the inhibition threshold for illegitimately reporting an actual number of their series may be lower compared to directly entering a fraudulent number (see Shalvi et al. 2011 for an empirical confirmation of this argument).

Furthermore, subjects were asked to enter their earnings in addition to their thrown numbers. This allowed testing as to whether they understood the rules of the game. The devaluation of the number six was on purpose. This potentially increased the propensity for norm violations. In most board games, the number "six" is a desirable outcome. Therefore, subjects may feel particularly frustrated if they threw a "six", which is normally the best outcome.

The dice experiment was conducted at the end of four unrelated other experiments. The subjects were recruited from ETH Zurich and University of Zurich during May 2009 and May 2010. The experiments were conducted using the software z-Tree (Fischbacher 2007). Thirty sessions were conducted, each of which with fourteen to sixteen subjects. In total, 466 subjects participated in the experiment. There were 63 % male subjects with an average age of 23 years (*sd* = 3.2). After the experiment, a small questionnaire was administered eliciting socio-demographic information such as income, education of the parents, religious affiliation and religiousness.

IV. Results

Is the honesty norm violated at all? The Null hypothesis states that each number is rolled with a probability of 1/6 (i.e. 16.7 %). Even in the first throw, there is an apparent discrepancy between what we see and what we would observe if everybody was honest (see Figure 1). The relative frequency of the highest payoff is 35 % in the first and 40 % in the second throw (averaged over all treatments). An 'honest six' (with zero earnings) is only reported by seven percent in the first and by six percent in the second throw. The expected probability of throwing twice a five is 1/36 (2.8 %). Despite of this, 20 % report this occurrence, which is over seven times as much as what we would observe in an honest population (Figure 2). However, subjects are more cautious with reporting double sixes, which happens in only one percent of the cases.

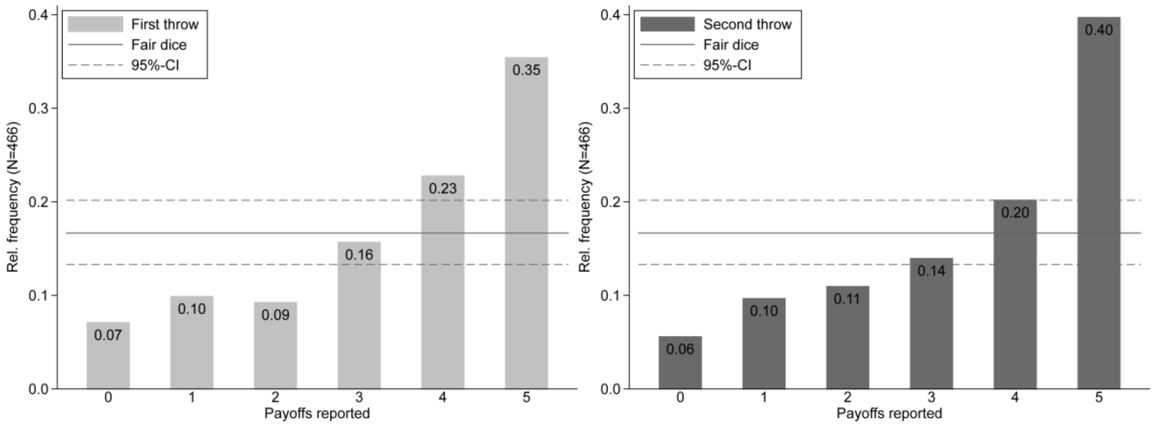

Figure 1: Distribution of reported numbers in the first and second throw (averaged over all treatments)

Interestingly, there is even fraud below the maximum. Apparently, people make compromises between their compliance with the honesty norm and their self-interest. Figure 2 shows the distribution of the sum of both throws. The number nine occurs in 16 % of the cases although its expectation value would be 5.6 %. Possibly, a combination of four and five is frequently reported, because it seems less suspicious than a double five.

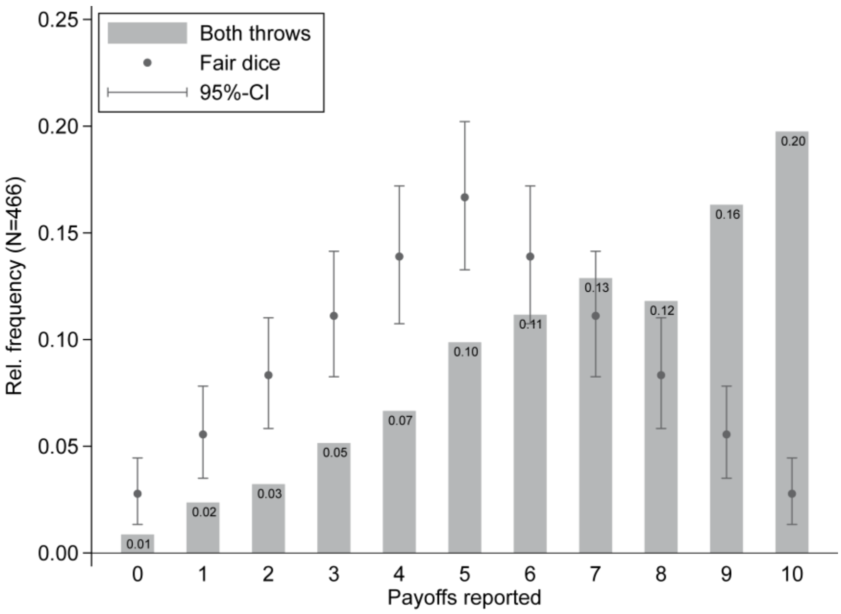

Figure 2: Distribution of reported cumulated payoffs for both throws (averaged over all treatments)

While the previous analyses demonstrate that a substantial fraction of the population violates the honesty norm and claim more money than they are entitled to, the question remains as to whether lying behavior is even more wide-spread if people are informed about the lying behavior of others. Hence, we compare the extent of norm violations across the three experimental treatments. Figure 3 shows the differences between first and second reported throws in the control condition (info 0), in the condition with information about the external large group (info 1) and in the condition with the information about the internal small group (info 2). A comparison between both information conditions and the control condition yields significant differences (ANOVA, $F_{2,29} = 4.90$; $p = 0.015$; see model 1 in the appendix for further details). Note that there is no significant difference between the experimental conditions (info 1 vs. info 2), suggesting that the kind of information feedback is less important than the fact of information feedback. In contrast to a comparison of the experimental conditions with the control condition, it is also possible to test whether the differences between the mean reported payoffs in the first and second throw are different from zero in all three treatment conditions. This is tested by a linear regression model without intercept (see model 2 in the Appendix). A joint test yields that the treatment differences are significantly different from zero (ANOVA: $F_{3,29} = 3,47$; $p = 0.029$). However, each separate treatment effect is not significantly different from zero. The p-values for the difference of the control condition is $p = 0.10$, for treatment condition 1 $p = 0.07$ and for treatment condition 2 $p = 0.14$. However, the effects of both treatment conditions are in line with the hypothesis and support the conclusions from the linear regression model without intercept, which yielded significant differences between both experimental conditions and the control.

Figure 4 shows that the rate of subjects who claim the highest payoff in the first and the second throw of the control condition is almost similar. However, this rate increases in both experimental conditions. Hence, the rate of norm violations increases if people are informed about the extent of norm violations in their own group or in an external group. We can confirm the statistical significance of this finding by logistic regression models, using the probability of a reported five as the outcome variable and the experimental condition as the predictor. (Note that confidence intervals in figure 4 are computed from this logistic regression.) We further substantiate our results by linear regression models, in which we compute the difference between first and second throws by experimental condition. This confirms that the average reported numbers are higher in both experimental treatments compared to the control treatment (see Table A1 in the appendix).

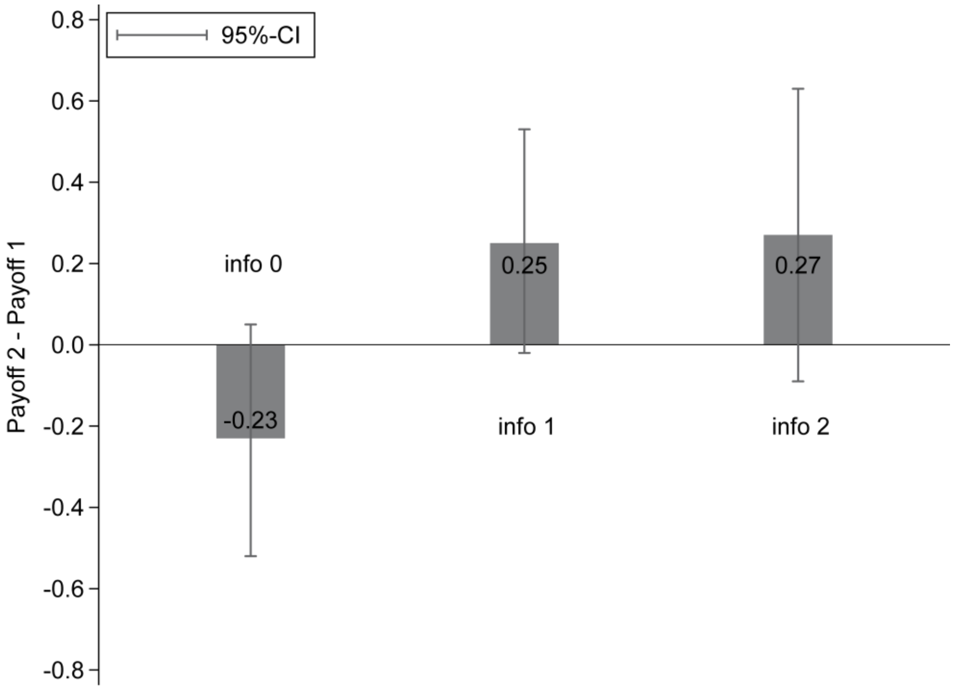

Figure 3: Difference between first and second reported throws by experimental condition

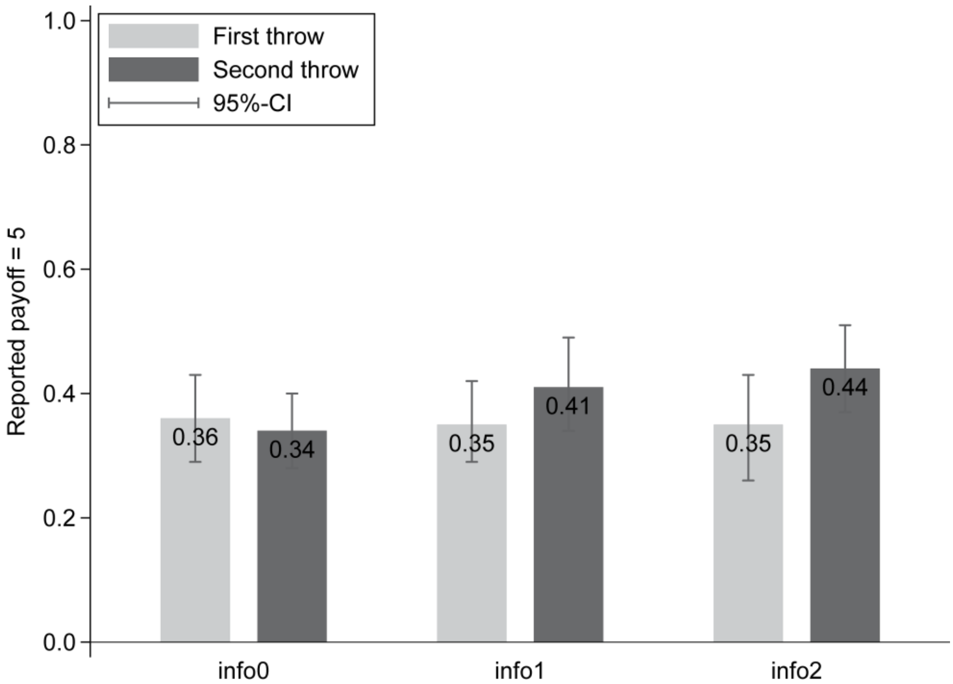

Figure 4: Fraction of reported maximum payoff (throwing a “five”) by experimental condition

Our second treatment condition (info 2) offers a more specific test of the ignorance hypothesis. In this condition, subjects have seen different rates of reported first payoffs, because information feedback was based on the specific sessions subjects were participating in.[4] Sessions varied substantially regarding the distribution of reported first payoffs. This variation may have partly been due to different propensities for lying and partly to pure randomness in throws. We can exploit session variability in lying as indicator for the extent of revealed lying. In some sessions, an almost even distribution indicated a small extent of lying, while in others, a strongly skewed distribution indicated a large extent of lying (see Figure 5 for the two most extreme sessions).

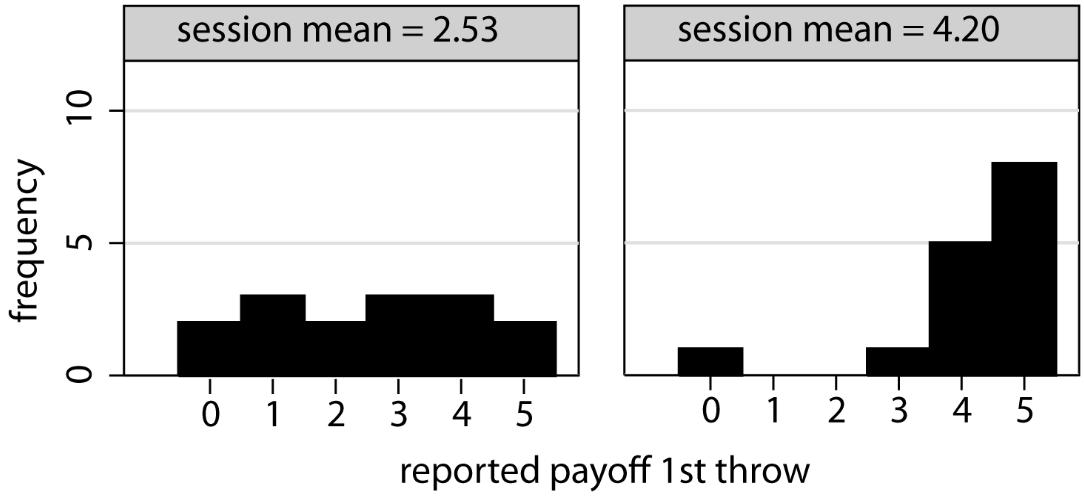

Figure 5: Illustration of the two sessions with the smallest and largest session average of reported first payoffs.

Our more specific ignorance hypothesis states that a larger extent of revealed lying in first throws triggers more subsequent lying in second throws. We used the session mean as indicator of revealed lying. We regressed session means of reported first payoffs on individuals' propensity to report a five in the second throw (Table 2, model 1) and on the

[4] The session data is structured as follows. The experimental sessions consisted of fourteen to sixteen subjects. In each session, all subjects were partitioned into the three treatment conditions so that each treatment consisted of four to six subjects within each session (except session 19, which only consisted of treatments one and two).

reported mean payoff in the second throw (Table 2, model 2). Both regressions only refer to data from information treatment two, where information about distributions varied.

The more specific ignorance hypothesis has to be rejected. The analysis in Table 2 reveals that session variability in information feedback regarding initial lying does not have statistically significant effects on subsequent lying. We conducted a number of additional robustness checks of the null finding (see the appendix for more details). Our robustness checks consisted of different operationalizations of revealed lying behavior, such as using the session skewness instead of session means. We also tried different statistical specifications of the hypotheses by taking different models, such as Poisson regressions. Furthermore, we ensured by simulation scenarios that the variability in sessions would have been sufficient to elicit significant findings and to make sure that our models are not affected by statistical artifacts like the so-called "regression to the mean". All these analyses indicate that the more specific ignorance hypothesis has to be rejected.

Table 2: Regression models quantifying whether a higher indicator of lying about first throws in sessions (measured by session means) increases the number of reported five have (1) and reported payoffs (2) in second throws

	(1) five in 2nd throw *only info 2 treatment*	(2) reported payoff 2nd throw *only info 2 treatment*
session mean 1st throw	-0.067 (-0.17)	0.029 (0.085)
intercept	-0.012 (-0.0089)	3.58** (3.01)
N	150	150

Notes: Model (1) reports a logistic regression model for the propensity to report a five in the 2nd throw explained by the mean payoff in the 1st throw in each session. Model (2) reports a linear regression model for the reported payoff in the 2nd throw explained by the mean payoff in the 1st throw in each session. Both models refer only to data from information treatment two, which is the only treatment with information feedback regarding the distribution of payoffs in the experimental session. Both regressions take clustered standard errors for sessions into account, t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Another interesting exploration is to estimate the extent of honesty and lying from the aggregate distributions of reported payoffs. It is possible to distinguish between two pure types of players; *moralists* and *liars,* based on the statistical computation of Fischbacher and Heusi (2008, p. 12 f.). Regardless of their actual entitlements, moralists report the truth and liars report the highest payoff.

The proportion of moralists can be estimated with respect to the number of individuals who report a zero payoff. One can extrapolate this to the whole population because it is plausible that somebody who is willing to report zero payoffs would truthfully report any higher payoff. Therefore, the full proportion of moralists can be estimated by multiplying the frequency of reported zero payoffs by six. The proportion of liars can be estimated with respect to the number of reported five's – those who claim the highest payoff. The difference between the expected percentage of five's for a fair die (1/6) and the empirically reported proportion of five's ("*empfive*") returns an estimate of liars in the population. In addition, this estimate has to be adjusted, because there are liars who actually threw a five, but would lie if they threw any lower number. Thus, the estimate of liars is corrected by the multiple 6/5. Therefore, the estimate of liars is (*empfive* - 1/6) * 6/5. Finally, there are actors who try to disguise their lies, who we call *deceivers*. These actors only partially increase their payoffs in order to disguise their lies. They report anything between two and four. We take a simple estimate of the frequency of this type by assuming that they represent the remaining population (100 % - moralists - liars = deceivers).

Figure 6 reports the estimated proportion of moralists, deceivers and liars in the second throw separately for each experimental treatment. The largest fraction, about half the population, consists of deceivers. There are fewer moralists (roughly 30 %). Liars are least frequent (roughly 20%). The figure also allows comparing the different distributions of lying, deceiving and honest behavior by treatments. There are about 10% - 20% less moralists and about 10% more deceivers and more liars in the experimental conditions with information feedback compared to the control condition without information feedback.[5]

[5] We mainly intend to report estimates of honesty, deception and lying by Figure 6. One statistical problem of estimating differences in types with regard to treatments is the low statistical power of estimates of moralists because they are only based on the relatively low number of reported zero payoffs. This also affects the estimates of deceivers, which are based on the remaining fraction after having accounted for moralists and liars. The estimates of liars are much more robust due to the much larger number of reported five's. This is why we do not discuss in detail significance tests with regard to differences in the proportion of moralists, deceivers and liars between the three treatments. Table A2 in the appendix reports respective bootstrap z-values and robustness checks with logistic regressions. The differences are not statistically significant; however, all effects have the expected direction.

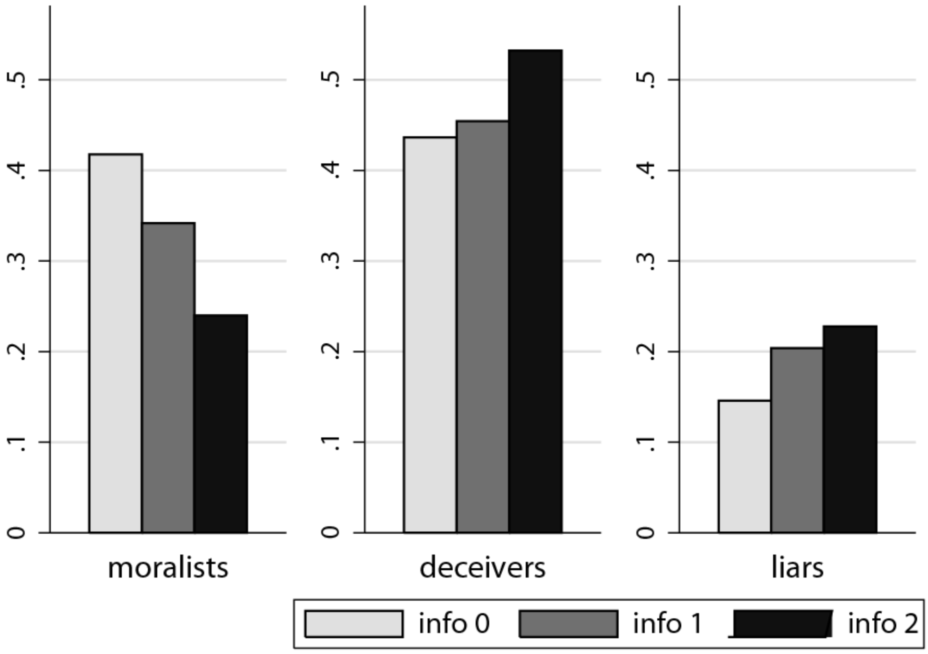

Figure 6: The proportion of different types in the second throw by experimental condition

Finally, we explored the socio-demographic determinants of lying. We estimated an ordered logit model, which predicts the reported number in the first throw; hence the "individual" propensity of lying without confounding social effects regarding the information of lying behavior of others. Participants with a higher monthly budget lie less, seemingly because they are less needy. Older persons lie more (see the model OLOGIT in Table A1 in the appendix). Protestants and Catholics are not significantly different from undenominational participants in their propensity to lie.

V. DISCUSSION

Our key point is to demonstrate the contagiousness of norm violations by the exclusive mechanism of conditional norm compliance. This mechanism is the "purest" kind of contagiousness, where actors simply imitate norm violations of others. Our experimental design excluded any punishment threats by generating absolute anonymity of norm violations. This excluded any explanation based on rational belief updates or cost-benefit analyses of getting caught and punished. Our findings therefore demonstrate that normative behavior implies "automatic", non-strategic decision-making when it comes to compliance with or violation of social norms. Our results suggest that merely information of norm violations of others is sufficient to trigger its spread.

Our experiment investigated whether lifting the veil of ignorance of the extent of others' norm violations triggers subsequently higher rates of norm violations (*ignorance hypothesis)*. This could be confirmed: Giving information about the extent of lying significantly increased subsequent lying. This effect was not due to learning, habituation or history effects since the effect did not occur in the control condition, which provided no information feedback. Further, this effect was robust to the specific kind of information feedback; either knowledge about norm violations of the own group or of a group of similar others.

The session analyses suggest that pure information feedback is sufficient to provoke subsequent cheating. It is apparently not the case that a greater extent of revealed cheating triggers even more cheating. This could mean that information feedback gives subjects the idea that others are cheating and this awareness is sufficient to trigger the erosion of the honesty norm. The actual extent of cheating seems to be less relevant. Further, it could be that subjects have difficulties to calculate the extent of cheating from the distribution of reported payoffs and therefore only take the fact into account that cheating does occur. A third alternative is that the effects are too small to be detected with 466 cases. However, the sample size is already quite large. While the sample size is large, our design yields inefficient estimates with the advantage of high anonymity and the disadvantage of statistical inefficiency. Subsequent studies could explore other designs yielding more efficient estimates.

In our study, we implicitly assumed that the rate of undetected norm violations is underestimated. This implies that underestimation stabilizes the norm. The larger this bias, the stronger the effect of ignorance on norm compliance. In contrast, if the extent of norm violations was overestimated, the reverse effect could occur; a preventive effect of knowledge. This could also be tested in prospective experiments.

The erosion of social norms is typically a gradual, subtle and slow process. Social norms do not disappear overnight. It is even more astonishing that the effects are detectable in a short experiment. Possibly, the honesty norm would further deteriorate if we allowed for a substantial continuation of norm violations and respective information feedback; a conjecture which could be tested in follow-up studies.

The conclusion that contagiousness of norm violations can exclusively operate by conditional norm compliance relates our findings not only to social psychology but also to experimental economics. Here, the recent notion of *conditional cooperation* describes the experimental finding that people condition their contributions to public goods on their beliefs of what they think others would contribute (Fischbacher et al. 2001). Individuals with cooperative intentions cease to cooperate if they become aware of sufficiently many freeriders (Gächter 2007). The finding of conditional cooperation in experimental economics overlaps with sociological and social psychological evidence showing that individuals align their behavior with the behavior of others (Goldstein, Cialdini and Griskevicius 2008; Schultz et al. 2007). While the economic concept of conditional cooperation refers to strategic reasoning based on reciprocity, the sociological concept of conditional norm compliance is more general and also applies to non-strategic, automatic imitation behavior.

Because our design excluded explanations based on strategic motives, our findings underline the non-strategic nature of normative behavior. In our experiment, there is no “rational” explanation of why actors condition their behavior on others, because liars in the group did not affect payoffs of other members. This distinguishes our findings from conditional cooperation in public goods experiments, where free-riders in the group lower the payoffs of all members. This gives sound evidence of the “non-rational”, imitative basis of normative behavior.

Acknowledgements:

We thank James A. Kitts, Ryan O. Murphy, Karl-Dieter Opp and Fabian Winter for valuable comments.

References

Allingham, M. G. & Sandmo, A. (1972), 'Income tax evasion: a theoretical analysis', *Journal of Public Economics* **1**(3-4), 323-338.

Banerjee, Abhijit V. 1992. "A Simple Model of Herd Behavior." *The Quarterly Journal of Economics* 107:797-817.

Becker, G. S. (1968), 'Crime and Punishment: An Economic Approach.', *Journal of Political Economy* **76**(2), 169-217.

Bullough, V. L. 1998. "Alfred Kinsey and the Kinsey Report. Historical Overview and Lasting Contributions." *Journal of Sex Research* 35: 127--131.

Bikhchandani, S., D. Hirshleifer, and I. Welch. 1992. "Learning from the behavior of others: Conformity, fads, and informational cascades." *Journal Of Economic Perspectives* 12:151-170.

Cialdini, R.B., R.R. Reno, and C.A. Kallgren. 1990. "A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places." *Journal of Personality and Social Psychology* 58:1015-1026.

Fischbacher, U., S. Gächter, and E. Fehr. 2001. "Are people conditionally cooperative? Evidence from a public goods experiment." *Economics Letters* 71(3):397-404.

Fischbacher, Urs. 2007. "Z-Tree. Zurich Toolbox for Ready-made Economic Experiments." Pp. 171-178 in *Experimental Economics*.

Fischbacher, U. & Heusi, F. (2008), 'Lies in Disguise. An experimental study on cheating', *TWI Research Paper Series*.

Friedman, T. L. (2011), “This Is Just the Start”, *The New York Times*, March 1, 2011 http://www.nytimes.com/2011/03/02/opinion/02friedman.html.

Gächter, Simon. 2007. "Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications." Pp. 19-50 in *Economics and Psychology. A Promising New Cross-Disciplinary Field*, edited by Bruno S. Frey and Alois Stutzer. Cambridge (MA): MIT Press.

Geserick, Rolf (1989): *40 Jahre Presse, Rundfunk und Kommunikationspolitik in der DDR*. München.

Gino, F., Ayal, S. and Ariely, D. (2009), 'Contagion and Differentiation in Unethical Behavior', *Psychological Science* **20**(3), 393.

Goldstein, Noah J., Robert B. Cialdini, and Vladas Griskevicius. 2008. "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels." Journal of Consumer Research 35(3):472-82.

Groeber, Patrick, and Heiko Rauhut. 2010. "Does ignorance promote norm compliance?" *Computational Mathematical Organization Theory* 16:1-28.

Keizer, Kees, Siegwart Lindenberg, and Linda Steg. 2008. "The Spreading of disorder." *Science* 322:1681-1685.

Kinsey, Alfred C. and Pomeroy, Wardell B. and Martin, Clyde E. 1948. *Sexual behavior in the human male*: Saunders.

Kinsey, Alfred C. and Pomeroy, Wardell B. and Martin, Clyde E. and Gebhard, Paul H. 1953. *Sexual behavior in the human female*: Saunders.

Marks, David. 1983., "Broadcasting Across the Wall: The Free Flow of Information Between East and West Germany", Journal of Communication 33(1), 46-55.

Popitz, Heinrich. 1968. *Uber die Präventivwirkung des Nichtwissens: Dunkelziffer, Norm und Strafe*. Tübingen: Mohr.

Sampson, R.J., and S.W. Raudenbush. 2004. "Seeing Disorder: Neighborhood Stigma and the Social Construction of 'Broken Windows.'." *Social Psychology Quarterly* 67:319--342.

Schultz, P. Wesley , Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2007. "The Constructive, Destructive, and Reconstructive Power of Social Norms." *Psychological Science* 18(5):429-34.

Seneca, Lucius A. 1928 [first 63]. *Moral Essays*. Translated by John W. Basore. The Loeb Classical Library. London: W. Heinemann, 1928-1935. 3 vols.: Volume I. http://www.stoics.com/seneca_essays_book_1.html#MERCY1

Shalvi, S.; Dana, J.; Handgraaf, M. & De Dreu, C. (2011), 'Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior', *Organizational Behavior and Human Decision Processes*.

Spiegel. 1980. "Vielköpfige Hydra". Der Spiegel, November 3, p. 42-46.

Thackeray, W. M. (1869), *On being fond out*, London, pp. 125-132.

Warner, Stanley L. 1965. "Randomized response: A survey technique for eliminating evasive answer bias." *Journal of the American Statistical Association* 60:63-69.

Wilson, James Q., and George L. Kelling. 1982. "The police and neighborhood safety: Broken windows." *Atlantic Monthly* 127:29-38.

Appendix A1: Regression models for ignorance hypothesis

Table A1: Regression models on the preventive effects of ignorance

	OLS 1 (payoff 2 – payoff 1)		*OLS 2* (payoff2 – payoff1)		*LOGIT* (payoff = 5)		*OLOGIT* (payoff 1)	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
experiment								
first throw					(ref.)			
control group (info 0)	(ref.)		-0.234	0.138	-0.054	0.167		
experimental group (info1)	0.487*	0.195	0.253	0.135	0.243	0.206		
experimental group (info2)	0.501*	0.189	0.267	0.177	0.360*	0.182		
religion								
undenominational							(ref.)	
protestant							0.274	0.236
catholic							0.027	0.314
other							0.975*	0.372
religiousness							-0.057	0.075
individual								
gender (male = 1)							0.148	0.176
age (in years)							0.053*	0.025
monthly budget (in 100 CHF)							-0.023*	0.011
years education father							0.015	0.044

years education mother							-0.053	0.037
Intercept	-0.234	0.138			-0.601*	0.105		
threshold 1							-2.099*	0.947
threshold 2							-1.039	0.920
threshold 3							-0.458	0.929
threshold 4							0.257	0.924
threshold 5							1.239	0.934
N_1	466		466		932		431	
N_2	30		30		30		30	
adj. R^2	0.01		0.01					
pseudo R^2					0.004		0.01	
χ^2					5.827		37.43	

Note: The table reports coefficients with clustered standard errors (* $p < 0.05$) for linear, logistic and ordered logistic regression models. Column 1 reports a linear regression for the difference between the first and the second reported payoff. Here, the control condition is the reference category so that the coefficients express the difference between the respective experimental and the control condition. Column 2 reports the same linear regression without an intercept. This means that the coefficients test whether the difference between the mean reported payoff difference between first and second throw is different from zero in the respective condition (control, info 1 and info 2). Column 3 reports a logistic regression model for the likelihood to report the maximum payoff of five. The model is specified such that the reference category is the first throw. This means that the coefficients express the difference between first and second throw for each experimental condition. A negative coefficient means that the likelihood to report a five was lower in the second throw than in the first throw. A positive coefficient expresses an increased likelihood to report a five in the second throw. Column 4 reports an ordered logit model for the first reported payoff. The experimental conditions are omitted so that the coefficients express the sociostructural determinants of lying.

Appendix A2: Specific ignorance hypothesis about session variability in info 2 condition

We performed a number of additional robustness checks of the finding that the specific ignorance hypothesis has to be rejected (Table 2 in the main text). First, we confirmed that our data is rich enough to allow for the detection of session effects. Figure A1 illustrates the session variability of reported first payoffs and compares it with simulation models of a fair die with a comparable number of observations. This demonstrates that there is considerable variation in reported payoffs in different sessions. Figure A2 shows that there is considerable variation of session means and further illustrates that the session means are much higher than one would expect from a fair die.

We also checked that the null findings are not due to too little variation in different numbers of reported five's in each session (see Figure A3). We further tried different operationalizations of our hypothesis. We analyzed the data on session level by performing a Poisson regression of the number of reported five's in the 1st throw in each session predicting the number of reported five's in the 2nd throw; yielding no significant effects. Other variants returned similar results as using the skewness of the distribution of reported first payoffs instead of the mean payoffs or the number of reported five's. We further complemented the empirical regression analyses with respective simulation models of a fair die with a large number of observations to make sure that our models are not affected by statistical artifacts like the so-called "regression to the mean". Finally, we compared the regression models with similar models for the data of the other two information treatments. All these additional analyses indicate that the specific ignorance hypothesis has to be rejected, i.e. that session variability of initial lying has no significant effects on subsequent lying.[6]

[6] The mentioned additional robustness analyses are available on request.

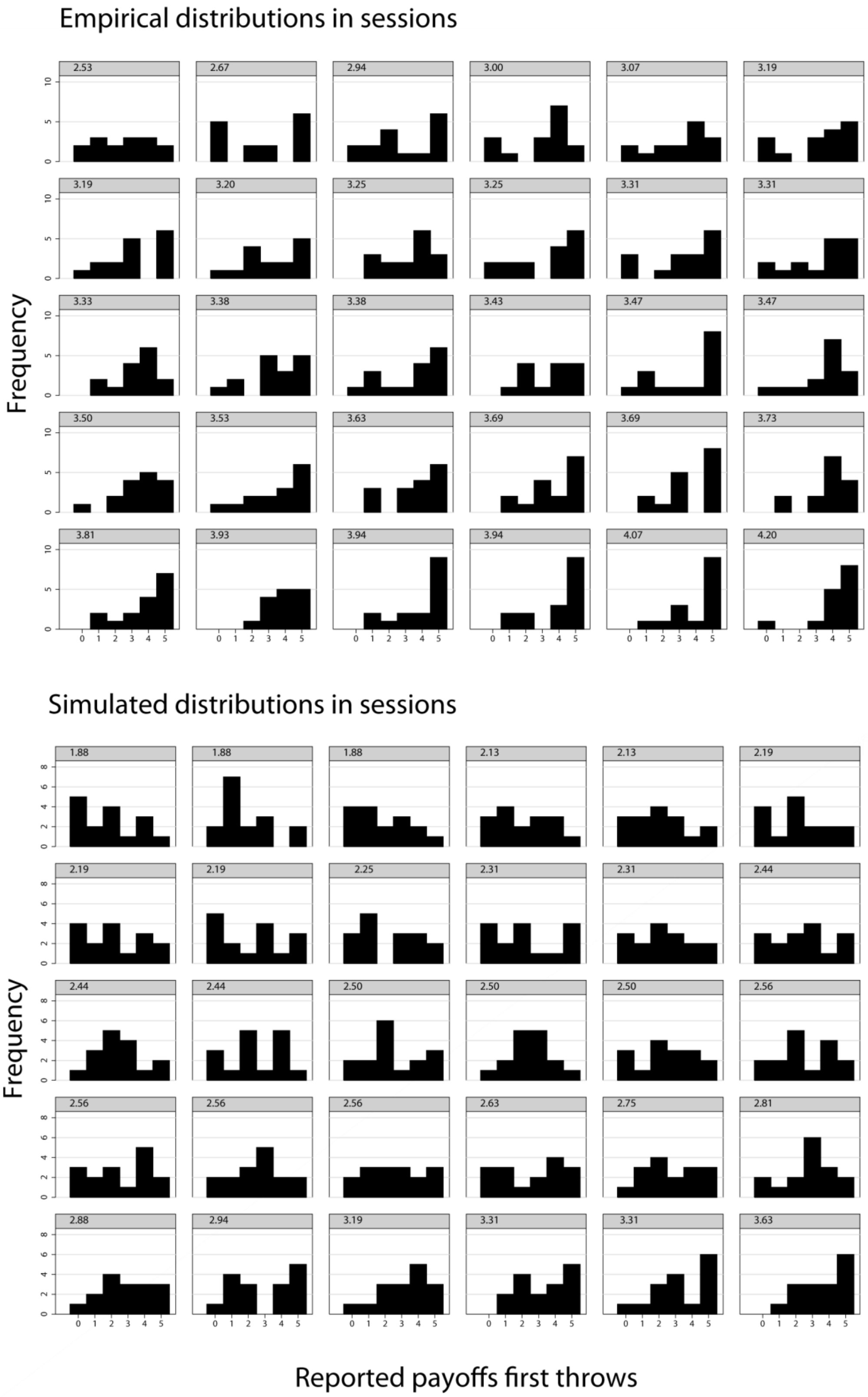

Figure A1: Distribution of payoffs in 1st throw by session, (A) sorted by mean session payoff. (B) Simulated distribution of payoffs in 1st throw sorted by mean session payoff (30 groups of 16 subjects, 480 observations).

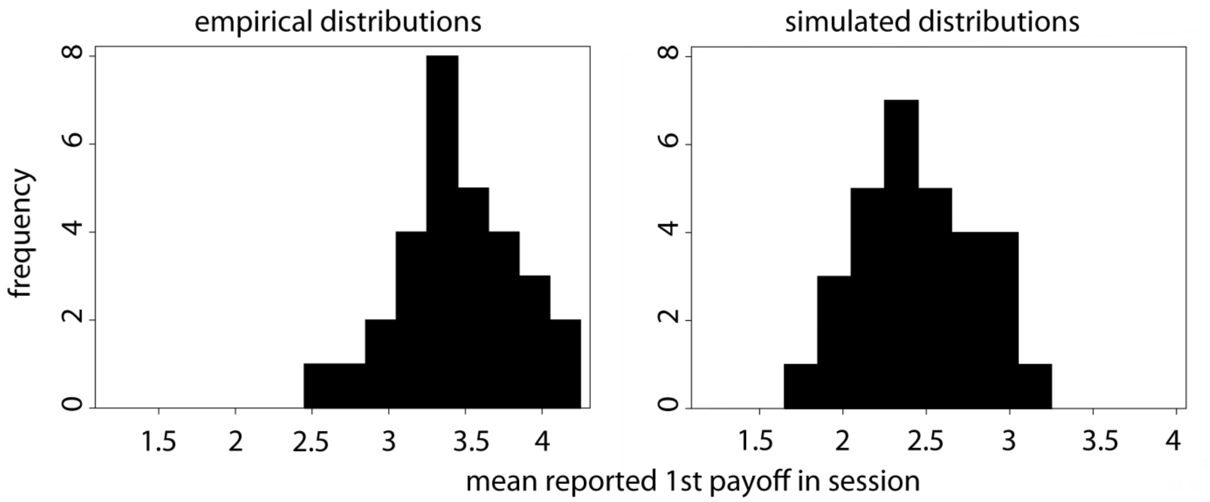

Figure A2: Distribution of arithmetic mean's for each session in (left) experimental data and (right) simulated data (30 groups of 16 subjects, 480 observations)

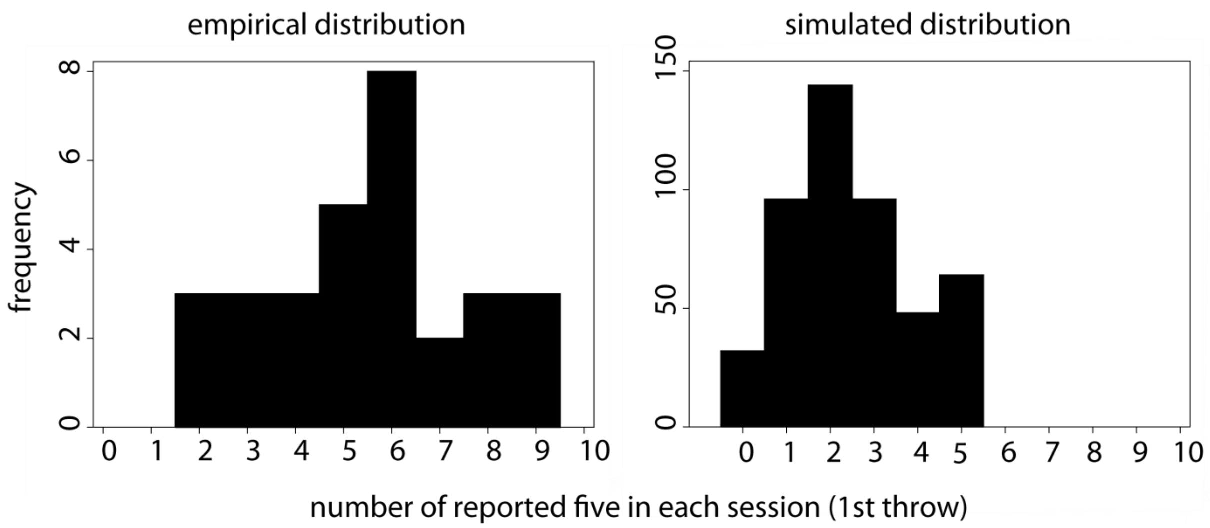

Figure A3: Distribution of number of five's in 1st throw for each session in (left) experimental data and (right) simulated data (30 groups of 16 subjects, 480 observations)

Appendix A3: Analysis of distributions of honesty types in different treatments

Table A2: Bootstrap z-values for the differences of the extent of moralists, deceivers and liars between each experimental treatment and the control group in the second throw.

	moralists *coefficient* *(bootstrap z-value)*	deceivers *coefficient* *(bootstrap z-value)*	Liars *Coefficient* *(bootstrap z-value)*
Difference info 2 - info 0	- 0.18 (-1.13)	0.096 (0.63)	0.082 (1.73)
Difference info 1 - info 0	- 0.076 (-0.46)	0.018 (0.11)	0.058 (1.27)

Notes: The bootstrapped differences are computed with 1000 replications. The estimated boostrapped differences for moralists and liars are essentially similar to respective logistic regression models which take the treatments as independent and the extent of reported five's or zero's in the second throw as dependent variables. Because the extent of deceivers is computed as rest category (100 % - moralists – liars), it is not possible to conduct a respective regression model as a robustness check for deceivers.

Appendix A4: Screenshots of experimental treatments

Now, you have the possibility to make some extra money.

You determine your additional payoff by casting the die in front of you on your desk. Please do not cast the die before you are asked to do so.

Caution! The first cast determines your payoff!

To check whether the die is working correctly, you may cast it several times. But only the first cast counts. Your payoff is calculated according to the table below.

Number on the die:	1	2	3	4	5	6
Payoff in CHF:	1.00	2.00	3.00	4.00	5.00	0.00

OK

Please cast the die now!

To check whether the die is working correctly, you may cast it several times. But **only the first cast counts**.

Enter the number on the die of your first cast and your payoff according to the table below in the corresponding fields.

Number on the die: 3

Payoff in CHF: 3

Number on the die:	1	2	3	4	5	6
Payoff in CHF:	1.00	2.00	3.00	4.00	5.00	0.00

OK

Figure A4: Instructions of the experiment (top) and the report of the 1st throw (bottom)

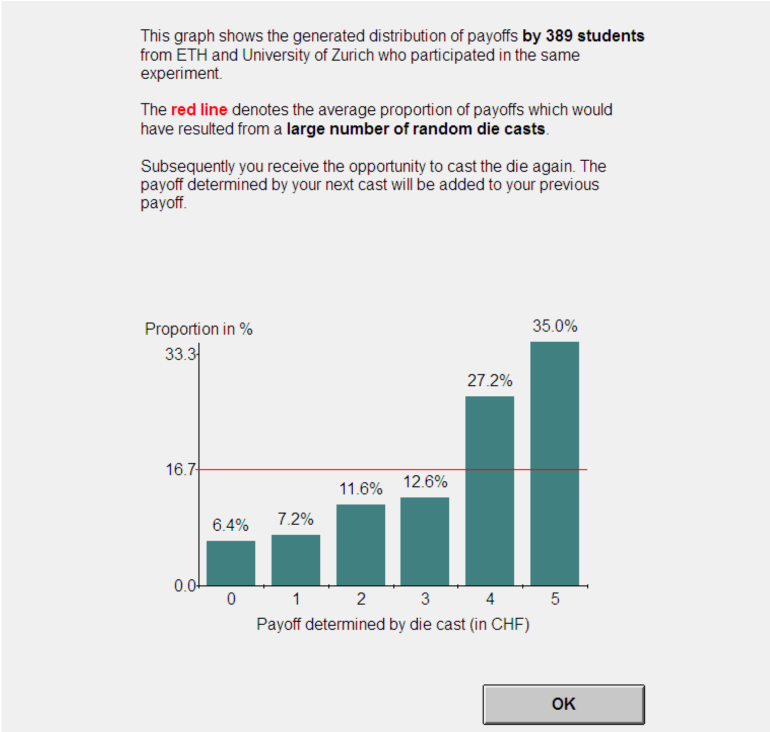

Figure A5: Screenshot of the 2nd experimental treatment (info 1) and the shown distribution of the experiment from Fischbacher und Heusi (2008)

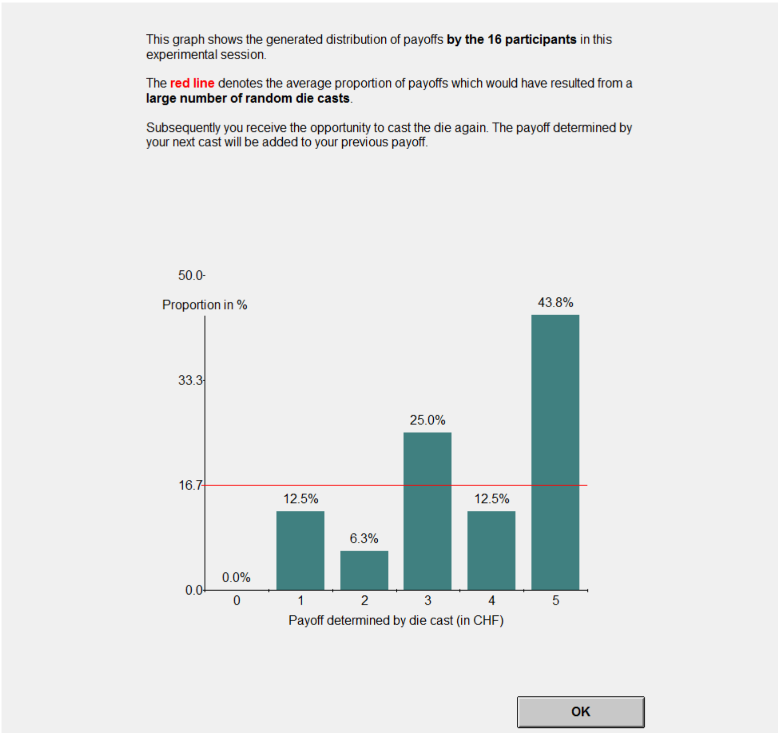

Figure A6: Screenshot of the 3rd experimental treatment (info 2) based on the distribution in session 17