games

MDPI

Article

What Goes Around, Comes Around: Experimental Evidence on Exposed Lies

Sarah Mörtenhuber [1], Andreas Nicklisch [2,*] and Kai-Uwe Schnapp [3]

1. Hauptverband der österreichischen Sozialversicherungsträger, Kundmanngasse 21, Postfach 600, 1031 Wien, Austria; sarah.moertenhuber@hvb.sozvers.at
2. Center for Economic Policy Research, HTW Chur, and DFG Research Group "Needs Based Justice and Distributive Procedures" (FOR 2104), Comercialstrasse 20, 7000 Chur, Switzerland
3. Institute of Political Science, University of Hamburg, and DFG Research Group "Needs Based Justice and Distributive Procedures" (FOR 2104), Allendeplatz 1, 20146 Hamburg, Germany; kai-uwe.schnapp@wiso.uni-hamburg.de

* Correspondence: andreas.nicklisch@htwchur.ch; Tel.: +41-81-2863736

Academic Editor: Ulrich Berger
Received: 22 June 2016; Accepted: 12 October 2016; Published: 20 October 2016

Abstract: We experimentally investigate the optimal way to handle the uncovering of a noble lie, that is, a lie that supposedly is in the best interest of a given community. For this purpose, we analyze a public good game with feedback to group members on the average contributions of the other group members. The computer program inflates the feedback and shows higher than real average contributions to the high contributors. As shown by earlier studies, the partial feedback inflation increases the total payoff of the public good as it avoids the feeling of being a sucker for above average contributors. The lie is then uncovered and we continue with different feedback modes on contributions, some inflated, some true. We find that players respond similarly to both feedback modes. However, with true feedback, initial contributions in the second stage are significantly higher than with inflated feedback.

Keywords: feedback inflation; public goods provision; noble lie; truth

JEL Classification: C91; D03; H41

1. Introduction

Folk wisdom has it that lying is endemic to politics. Citizens of any state are likely to say that they have encountered political lies off and on in all levels of government. The academic debate picks up on politicians telling lies in democratic systems and there are opposing positions about the legitimacy of such lies. Some claim that a political lie can help to actually attain certain public goods and that they are therefore justifiable (e.g., [1]). One of the most prominent contemporary examples for a noble lie is the Good Friday Agreement between Sinn Fein and the IRA on one side and the Ulster Unionists on the other side of the Northern Irish conflict in April 1998 [1]. Those defending these kinds of political lies call them "noble lies" with reference to Plato. Those opposing such views argue that lying is so

[1] Dixon ([1]) argues that the agreement was only possible because of political lying. Elites on both sides of the conflict and the British government had lied to their constituencies about the progress of the negotiations. According to Dixon's argument, by lying, the parties wanted to foster the willingness of the people in their own camp to agree to concessions that would not have been agreed to without the firm belief that the other side had already made such concessions in the negotiations. Although we will analyze lying and its aftermath in a different domain, namely the attempt to foster voluntary cooperation, the struggle for the Good Friday Agreement documents the potentially benevolent effect of political lies for societies as a whole.

foreign to democracy that it does nothing but undermine democracy itself [2], even in cases in which it does produce welfare benefits.

In our paper, we consider another important, but poorly developed aspect of political lying: we look at the long-term consequences of noble lies for the production of public goods with a focus on those (highly likely) cases in which the noble lie is uncovered. For this purpose, we analyze the results of a series of public good experiments in which we concentrate on the "aftermath" of the noble lie. More specifically, we ask whether it is beneficial in terms of efficiency to continue telling potentially reassuring lies or to provide the truth once a noble lie has been revealed. As such, we complement Gneezy et al. [3] who discuss the long-term consequences of lying in a sender receiver game. They show that the frequency of lying increases with experience when the gains of lying are high. In our design, continuation of lying may enhance or threaten the provision of a public good.

Specifically, participants in our experiment play two sequences of twenty rounds of a voluntary contribution mechanism game (VCM). In the first sequence, after every round, players receive feedback on the group's average contribution; the below average and average contributors receive correct information, and the above average contributors receive inflated feedback. The intention behind manipulating the information on the average contributions is to provide an incentive for the above average contributors to keep up their above average contributions in the next round as well. This would lead to the group gaining as a whole at the expense of the cooperative players [2].

There are three treatment variations in the second sequence: (1) BASELINE—the arrangement of the first sequence stays unchanged and players do not know that the feedback is partly inflated; (2) LIE—between sequences, the participants are told that their feedback in sequence 1 has been manipulated and they continue to receive feedback inflated in the same way during the second sequence. That means the noble lie is continuously applied, but confidence of the players in the accuracy of the feedback mechanism is (again) seriously unsettled by the revelation of the true feedback mechanism; (3) TRUTH—between sequences, the participants are told that their feedback has been inflated in sequence 1. They are then told that in the second sequence they will receive non-manipulated feedback. Cooperation rates are compared between all three treatments.

Treatment variations (2) and (3) represent two extreme communication strategies for the aftermath of the lie. One may argue that people—after being informed that they were lied to—appreciate this efficiency enhancing manipulation as they understand that they all profit from it (e.g., Hoffmann et al. [8]). Following this argument, the continuation of the lie has positive implications for an individuals' willingness to cooperate in comparison to telling the truth, since people anticipate the stimulating effect of the lie. On the other hand, it could be that people appreciate unambiguous information in comparison to potentially wrong messages even if this information indicates an income loss: people may interpret the potentially wrong message in exactly the same fashion, namely as information indicating an income loss, while the lie adds additional ambiguity, for instance, about the amount that is lost. Overall, it is unclear per se whether one of the two cases (i.e., telling lies or providing the truth) leads to more efficient outcomes. That is why we test them experimentally.

Our results from the second sequence of the VCM show significantly lower average contributions in LIE than in BASELINE. In contrast, the average cooperation rate in TRUTH falls between the ones from the other two treatments. Differences between TRUTH and the other treatments are not significant. We identify two effects leading to this finding: if the manipulation is revealed, contribution rates decline approximately at a similar rate independently of the type of feedback. However, the

---

[2] We would like to stress that we respect the convention in experimental economics not to deceive subjects by informing subjects prior to the start of the experiment that the displayed information on the others' average contribution might deviate from the actual value. In the words of Hey [4], we provide subjects with potentially wrong information, but tell them in advance that this can happen during the experiment. Following arguments by Hertwig and Ortmann [5], we claim that this is not deception, since deception would have occurred only if participants, after being completely debriefed, had perceived themselves as being misled. This is not the case due to our prior information. For further discussions concerning the effect of deception in experiments see, e.g., [5–7].

potentially manipulated feedback in LIE yields significantly lower levels of initial contributions than truthful feedback. Thus, playing another sequence with potentially manipulated feedback leads to lower welfare (in the sense of accumulated payoffs) across the entire second sequence than the inconvenient truth.

Our experimental results complement a study by Hoffmann et al. [8] who compare the welfare implications of different types of noble lies against a setting without lies [3]. They show that a noble lie to all participants does not travel far, while lying selectively to very cooperative players significantly improves the social welfare compared to the setting without the lie—even if the lie's nature is revealed between sequences. Both Hoffmann et al.'s and our paper follow a stream of literature analyzing moral concerns in lying: Gneezy [11] shows in his seminal paper that people trade off material payoffs against moral scruples in sender–receiver games. In the same context, Sanchez-Pages and Vorsatz [12] find a positive correlation between lying and punishing liars.

Serra-Garcia et al. explore the relation between lying and communication within repeated public good games. Serra-Garcia et al. [13] find that precise communication leads to frequent lying, while lying is less frequent when vague communication is allowed. In turn, communication receivers do not seem to realize that vague messages indicate truthful information. Serra-Garcia et al. [14] show that communication content that relates to promise components leads to less lying and less free riding. Related to this issue, Engel et al. [15] and Irlenbusch and Rilke [16] test the effect of selective feedback information in public good games. The latter paper shows that only if the selection criteria for feedback is not disclosed to the participants can samples with very cooperative players boost cooperation, while samples with very uncooperative players hamper it.

Building on this rich literature, we analyze the "best behavior" in the aftermath of a noble lie: we ask for the welfare effects of different feedback mechanisms after the noble lie has been exposed. In other words, we test whether truthful or inflated feedback yields superior payoffs in a second sequence of public good games following a first sequence of public good games with inflated feedback (the noble lie).

Our paper is organized as follows: we first discuss the details of our experimental game as well as the experimental procedure and formulate behavioral expectations. We then describe and analyze the results of the experiment and finally conclude with an answer to our questions and an outlook on future research concerning lying in politics.

2. The Experimental Design

2.1. The Game

To analyze our research question, we use the linear standard VCM as an experimental tool. This design has widely been tested experimentally by biologists, economists and political scientists (see Zelmer [17] and Chaudhuri [18] for surveys). The VCM allows us to investigate cooperation behavior in response to different feedback regimes. It is not tailored to a specific case or application but serves as a framework that incorporates important features common to many different situations.

All participants play two sequences of twenty periods of the following stage game in groups of four: at the beginning of each period, each player receives an endowment of 20 ECU (experimental currency units). Players simultaneously but covertly decide how many ECU they contribute to the public good, g_i, with $g_i \in \{0, 1, 2, ..., 20\}$. Each ECU contributed to the public good is multiplied by 1.6 and then divided among the four players. Thus, each ECU contributed yields a profit of 0.4 ECU (the marginal per capita return) to each player including i. Each ECU not contributed to the public good is privately kept and yields a benefit of 1 ECU to player i only. Therefore, player i's individual payoff

---

[3] Earlier studies by Marwell and Ames [9] and Weimann [10] also deal with manipulated feedback and find no significant effect. However, both studies provide feedback in a different format and use percentages of the total contributions instead of absolute numbers, which may trigger different responses.

function is $\pi_i = 20 - g_i + 0.4 \sum_{j=1}^{4} g_j$. All parameters are common knowledge for all participants of the experiment. Throughout the entire experiment, players remain in the same group of four players, both within each sequence, and across the two sequences (i.e., from period 1 to period 40).

2.2. Procedure

An experimental session proceeds as follows: once all subjects are seated, written instructions on sequence 1 are handed to them before the experimenter reads them out loud. Subjects are then given the opportunity to ask questions in private. Before the experiment starts, subjects answer a set of questions to test their understanding of the instructions. They are informed that there will be two sequences but no details on sequence 2 are given at the start of the experiment. At the start of sequence 2, subjects receive a new set of instructions. After completing sequence 1 and 2, subjects answer a short socio-demographic questionnaire. Finally, subjects receive profits from one randomly determined period of each sequence [4].

In the first sequence, we implement the following feedback mechanism at the end of each period: participants receive feedback on the average contribution of the other three group members $\tilde{g}_{-i}$ (rounded to the next full integer). Subjects contributing more than average receive inflated information: their feedback shows the amount of their own contribution as being the average (i.e., $\tilde{g}_{-i} \equiv g_i$). All other subjects are shown the actual average contribution of the other group members [5]. Notice again that we clarify at the beginning of the experiment that the feedback they receive during the experiment on group averages may deviate from the true average.

Thus, players contributing more than average are likely to be too optimistic in regards to the cooperativeness of the other group members. As a result, these subjects keep up their high contribution level. This increases the total amount of income for the group and may even foster higher contributions of the other players. The latter is possible because the increased true average is shown to the below and at average contributors, thus providing an incentive for those players to increase their contributions in the next round as well. This group welfare increasing effect of the above described inflated feedback is the reason why we call this manipulation a "noble lie".

In the second sequence, we introduce three treatment variations of the game. In the BASELINE condition, prior to the start of sequence 2, participants are shown the (displayed) average contribution of all other participants across the previous twenty periods. Furthermore, they are informed that they will play another twenty periods of the same game. Hence, the noble lie is not fully revealed. In the LIE and TRUTH condition, we reveal the exact feedback mechanism to all participants after sequence 1. Participants are informed that they received correct feedback if they contributed at or below average, and that feedback did equal their own contribution if they contributed above average. In addition, participants are informed about the actual average contribution $\tilde{g}^{A}_{-i}$ and the displayed average contribution $\tilde{g}^{D}_{-i}$ of all other participants across the previous twenty periods.

In the LIE condition, the conditionally inflated feedback mechanism is implemented in the following twenty periods again. However, now players know that receiving feedback equaling their own contribution could mean either of two things: (a) their contribution is exactly at the group average or (b) it is above the group average and they do not know the true value of this average. Finally, in the TRUTH condition, players are provided with correct feedback in sequence 2 and they are aware of this.

At the end of the second sequence, we inform participants in all treatment conditions about the true nature of the feedback provided in the two sequences, and their accumulated payoffs.

---

4 English translations of the experimental instructions are enclosed in Appendix A.

5 The Good Friday agreement serves as a good real-world example: Constituencies of both sides were "tricked" into an agreement through "inflated" private feedback by their elites regarding the negotiation's progress, while the public was (truthfully) informed that nothing had happened so far.

2.3. Predictions

Since subjects lose 0.6 ECU per ECU contributed, the individually optimal strategy is to keep the whole endowment—independently of the exact feedback participants receive. The collectively optimal strategy, however, is to contribute fully, since all contributed ECU's are multiplied by 1.6 before they are divided up among the players who then receive 0.4 ECU each. The rationale of individual and social optimal strategies remains unchanged if the game is played repeatedly over a finite number of periods. As a consequence, individual contributions to the public good reflect players' voluntary cooperation.

As shown by Hoffmann et al. [8], partially inflated feedback fosters overall cooperativeness substantially. Therefore, we expect higher contributions in the second sequence of BASELINE than in the other two treatments. The crucial question is: how do participants respond to the information that feedback was inflated in the first sequence? That is, how do subjects respond contribution-wise to potentially wrong feedback $\tilde{g}_{-i} = g_i$ in the second sequence of LIE compared to the corresponding correct but inconvenient feedback $\tilde{g}_{-i} < g_i$ in TRUTH. We call this latter feedback "inconvenient" because it informs the player that her fellow players are less cooperative than she is.

Suppose now that players are loss averse in the sense that they maximize their contributions as long as they receive a payoff from the public good at least equivalent to their contribution to the public good [6]. That is, instead of maximizing utility $\max u_i(g_i)$ (in the case of narrow self-interest) or $\max u_i(g_i, g_{-i})$ (in case of social utility), subjects simply $\max g_i$ subject to $g_i \leq 3 \cdot 0.4\tilde{g}_{-i} + 0.4g_i$. For the symmetric equilibrium, if all four subjects follow this strategy, this would obviously lead to full contributions by all subjects. However, if subjects hold diverse beliefs regarding contributions (and perhaps some subjects do not follow this strategy), rich dynamics may unfold.

In TRUTH, this leads to a decrease in contributions as long as

$$0.6g_i > 3 \cdot 0.4\tilde{g}_{-i} \Leftrightarrow \tilde{g}_{-i}/g_i < 0.5, \tag{1}$$

and to stable contributions otherwise.

In turn, players who contribute more than the average in LIE cannot follow the same rationale, since they are not informed about the true $\tilde{g}_{-i}$ (and are aware of this in the second sequence). Here, we assume that they form a belief regarding the true average contribution as a fraction δ ($0 \leq \delta \leq 1$) of their own contribution g_i. Therefore, in LIE Equation (1) changes to

$$0.6g_i > 3 \cdot 0.4\delta g_i \Leftrightarrow \delta < 0.5. \tag{2}$$

As mentioned above, we inform participants about the actual average contribution and the displayed average contribution of all other participants during the first twenty periods between the first and the second sequence. Therefore, the ratio $\tilde{g}^A_{-i}/\tilde{g}^D_{-i}$ seems to be a good proxy for δ.

Importantly, Equation (2) is independent from the actual contribution level. Therefore, we do not presume a gradual adaptation of contributions as predicted for TRUTH. Rather than adapting their contributions based on those of the others in the last period, contributors have to rely on their beliefs about the LIE condition. Hence, we assume that cooperative participants who usually contribute above the average do so more "cautiously" right from the start of the second sequence (i.e., after being informed about the true nature of feedback inflation). For this reason, we hypothesize

Hypothesis *Average contributions in* LIE *drop initially in the second sequence, while in* TRUTH *they decrease gradually in the course of the second sequence. As a consequence, total payoffs across the second sequence are higher in* TRUTH *than in* LIE.

---

[6] Fischbacher and Gächter [19] analyze the dynamics of public good contributions in greater detail.

3. Results

All in all, we ran ten sessions with 196 participants who were matched randomly into groups of four, totaling 49 groups; each subject participated in only one treatment condition. One group served as an independent observation and we collected 16 observations in BASELINE, 16 in LIE, and 17 in TRUTH [7]. The experiment was run in June and July 2013 as well as in February 2015 in the experimental laboratory of the Faculty of Business, Economics and Social Sciences at the University of Hamburg. The University of Hamburg is a large public university with approximately 40'000 students in the center of Hamburg, Germany. The experiment was programmed using the software package zTree [20] (version 3.5.1, Fischbacher and Schmid, Kreuzlingen & Zurich, Switzerland), and hroot [21] (version 2.0, Bock, Baetge and Nicklisch, Hamburg, Germany) was used for recruiting. Most participants were students (two non-students) with different academic backgrounds including economics; 58% were women, and the median age was 24. Sessions typically lasted 60 min and the average payoff was 11.51 Euros. Subjects received cash payments immediately after the session was over.

Let us start with the average contributions in the first and the second sequence of the experiment. Figure 1 shows the evolution of average contributions by periods and across treatment conditions. Notice that, in addition, it depicts the evolution of average contributions for the sequences of twenty rounds of a VCM without any feedback inflation (NULL). The comparison with our treatment conditions allows us to assess the overall benefits of applying a noble lie at all [8]. As expected, there are no significant differences between our three treatments in the first sequence. However, the development contrasts with the typical evolution of voluntary cooperation in a public good game where contributions constantly decrease over time (compare the NULL treatment). In our experiments, contributions remain stable at a level of about 11 out of 20 ECU with mean contributions at 10.6 in BASELINE, 11.2 in LIE, and 11.0 in TRUTH. Comparing the contributions between our treatment conditions, both for the first period and across all periods of the first sequence, there are no significant differences [9].

Interestingly, not all players are equally likely affected by the noble lie. That is, while there are about 22 percent of players who never or almost never receive inflated feedback (at most in two out of 20 periods), there are 25 percent of players who always or almost always see inflated feedback (at least in 18 out of 20 periods) [10]. Thus, there seems to be a sub-population of stable cooperators who (almost always) see false feedback in the first sequence. We will pay attention to this heterogeneity regarding the exposure to the royal lie below when we analyze how players in LIE and TRUTH choose their contribution in period 21 (i.e., immediately after they are informed about the exact feedback mechanism).

Turning to the second sequence of the experiment, the development of cooperation rates differs substantially between treatments after the feedback mechanism is revealed. It seems that players' confidence in the cooperation rate is unsettled, so that contributions decrease more rapidly in LIE and TRUTH in comparison to BASELINE. Mean cooperation rates across the entire second sequence

---

[7] Please note: For an earlier version of this paper we tested a fourth treatment (WARNING) where at or above average contributors received inflated feedback along with an additional signal that this feedback was wrong. We introduced this treatment to explore the contribution-wise difference between receiving feedback likely to be inflated, but not knowing the degree of inflation (LIE) and receiving obviously inflated feedback, but not knowing the degree of inflation (WARNING). Since results were very noisy and difficult to interpret (average contributions in WARNING were similar to TRUTH, while there was a large variance), we decided not to include those results in this paper. However, we are happy to share more information about these treatments upon request.

[8] The experiments—overall 48 subjects in 12 independent groups—were run at another German University under similar conditions (similar instructions, similar computer program, similar subject pool). Data was thankfully provided by Hoffmann et al. [8].

[9] $p > 0.5$ for all comparisons; we test group averages using two-sided Wilcoxon–Mann–Whitney rank sum tests.

[10] The remaining roughly 50 percent of players are rather uniformly distributed across the number of periods they see inflated feedback.

are 10.7 in BASELINE, 6.6 in LIE, and 8.1 in TRUTH [11]. The difference between LIE and BASELINE is significant ($p = 0.003$, two-sided Wilcoxon–Mann–Whitney rank sum test), while the differences between TRUTH and BASELINE ($p = 0.11$, two-sided test) and between LIE and TRUTH are insignificant ($p = 0.31$, two-sided test).

Despite the constantly decreasing contributions after the feedback mechanism is revealed in LIE and TRUTH, there is an overall gain from the introduction of the noble lie for players in comparison to NULL: average contributions across both sequences are 8.9 in LIE, 9.5 in TRUTH, and 5.9 in NULL. Differences among average contributions between LIE and NULL, and between TRUTH and NULL are significant [12].

As one sees in Figure 1, the difference between LIE on the one hand and TRUTH as well as BASELINE on the other is particularly pronounced in the first period of the second sequence. Here, mean contributions are 8.7 in LIE, 11.7 in BASELINE, and 11 in TRUTH; the differences between LIE and BASELINE (TRUTH) is (weakly) significant ($p = 0.03$ and $p = 0.08$, respectively, two-sided tests), while the latter two differ insignificantly ($p = 0.55$). Thus in line with our hypothesis, the initial contributions in LIE seem to be substantially influenced by the inflated feedback. The resulting difference between contributions in LIE and the other two conditions remains recognizable—though partly insignificant—until the end of the second sequence.

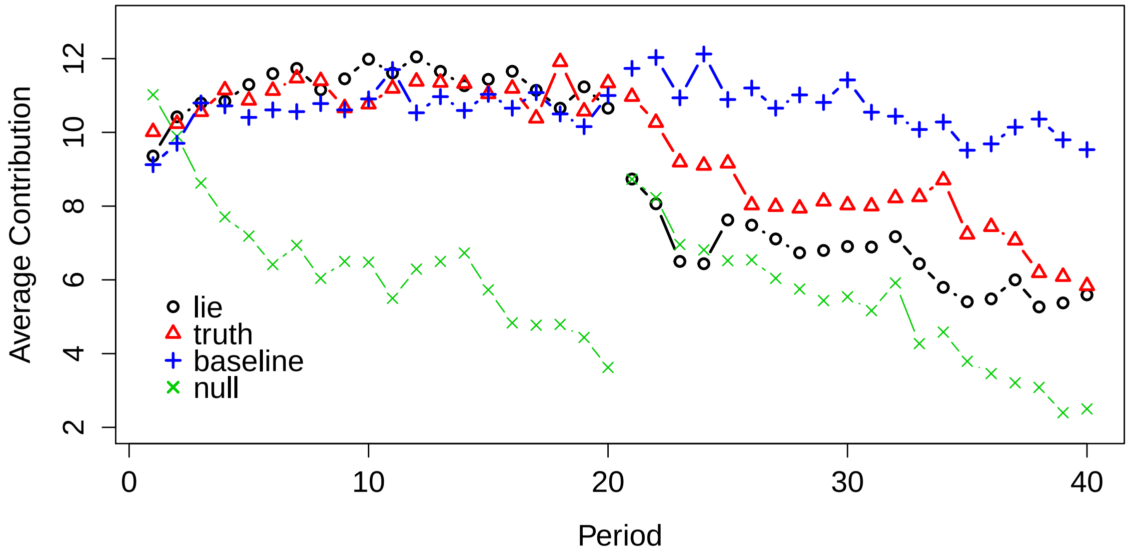

Figure 1. Average contributions by periods across treatments.

Result 1. *In sequence 2, initial as well as average contributions across the entire sequence are lower under* LIE *than under* BASELINE. *However, only initial contributions are lower under* LIE *than under* TRUTH, *while average contributions across the entire sequence do not differ significantly.*

We hypothesized earlier that initial contributions at the beginning of the second sequence drop more rigorously in LIE because informed adaptation of contributions in the course of the sequence is far more difficult for cooperative participants in this treatment than in TRUTH. To test this claim, we use the average ratio of feedback inflation $\tilde{g}^{A}_{-i}/\tilde{g}^{D}_{-i}$ they experienced during the first sequence (and

---

[11] As such, our results do not replicate the findings of Hoffmann et al. [8] who find decreasing contributions in the second sequence of their LIE-type treatment, but at a substantially higher overall level. We think it is plausible to explain that disparity with the difference in information provided on the kind of feedback manipulation between sequences: Hoffmann et al. [8] inform players that they see "at least their own contribution as the group average," while we state explicitly that they receive correct signals once they contribute at most the average, and that they receive their own contribution as signals otherwise.

[12] $p = 0.02$ and $p = 0.007$, respectively, two-sided Wilcoxon–Mann–Whitney rank sum tests.

which they learn at the end of the first sequence) as an indicator for their degree of cooperativeness. The smaller this ratio is, the more cooperative they were (relative to other group members).

In turn, we compute the drop in initial contributions as the difference between the contribution with which they start the sequence after being informed precisely about the feedback inflation (i.e., the g_i in period 21) and the contribution with which participants "normally" start a sequence of public goods games (i.e., the g_i in period 1). The premise of our hypothesis holds if we find a positive relation between ratio of feedback inflation in LIE and a zero or at least a weaker relation in TRUTH.

Figure 2 illustrates a slightly positive relation between the ratio of feedback inflation and the change in initial contributions in both treatments. Moreover, it seems that the relation is more pronounced in LIE than in TRUST. In other words, the importance of the observed accuracy of feedback in the first sequence for the initial contribution in the second sequence is substantially higher in LIE than in TRUTH [13].

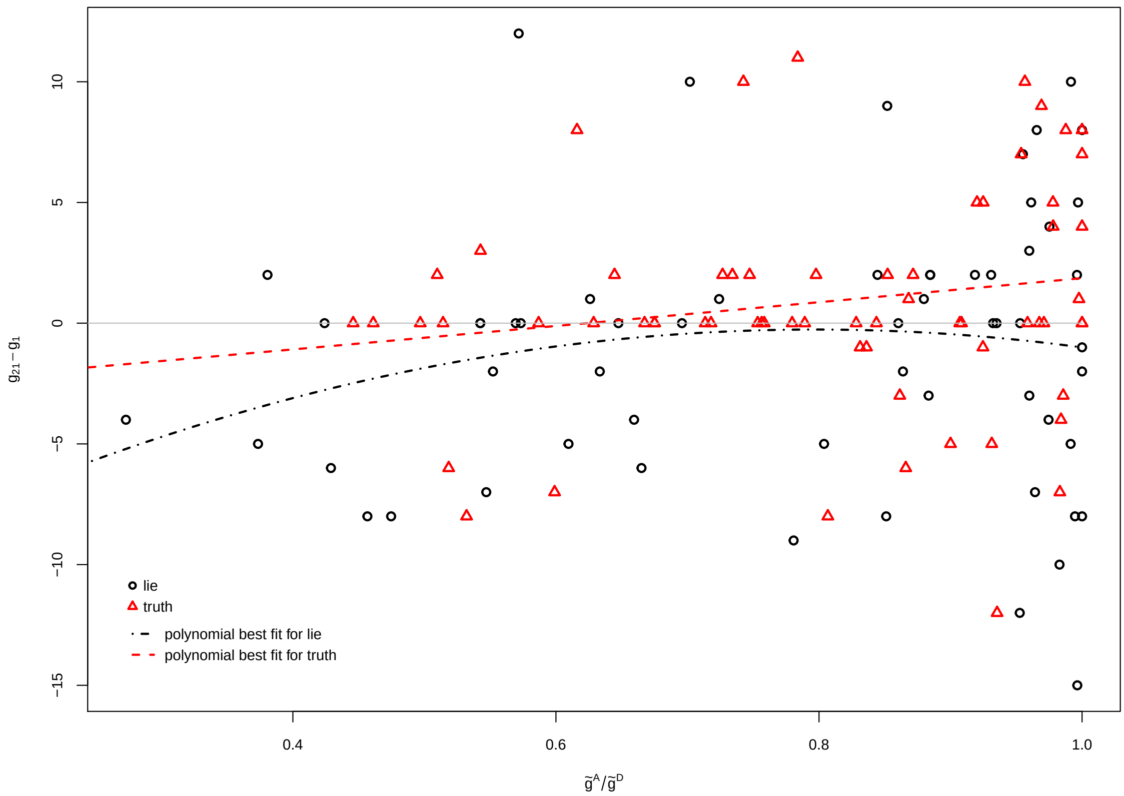

Figure 2. Differences between initial contributions in the second and first sequence related to $\tilde{g}^A_{-i}/\tilde{g}^D_{-i}$ along polynomial best-fit-curves in LIE and TRUTH.

We would like to stress that there is another insight from Figure 2: the dashed-dotted black (dashed red) line shows the polynomial best fit for $g_i^{21} - g_i^1 = \alpha_0 + \alpha_1 \tilde{g}^A_{-i}/\tilde{g}^D_{-i} + \alpha_2(\tilde{g}^A_{-i}/\tilde{g}^D_{-i})^2$ in LIE (TRUTH). This best fit is negative for above average contributors (with $\tilde{g}^A_{-i}/\tilde{g}^D_{-i} \ll 1$) and positive for below average contributors (with $\tilde{g}^A_{-i}/\tilde{g}^D_{-i} \approx 1$) in TRUTH implying that initial contributions for both groups converge in the second sequence on average. However, the best fit is inverted u-shaped and negative for all contributors in LIE, implying that all players decrease their contributions on average.

[13] To prove this point, we run a linear regression with $g_i^{21} - g_i^1$ from LIE and TRUTH as a dependent variable and $\tilde{g}^A_{-i}/\tilde{g}^D_{-i}$, the dummy variable *truth* (which equals one for observations from TRUTH and zero otherwise), and the interaction term *truth* $\times\, \tilde{g}^A_{-i}/\tilde{g}^D_{-i}$ as independent variables. The estimated coefficient for $\tilde{g}^A_{-i}/\tilde{g}^D_{-i}$ is significantly positive (3.98, standard error is 1.599), while for the interaction term, it is significantly negative (-3.13, standard error is 0.945). That is, the influence of the observed feedback on the change in initial contributions is significantly weaker in TRUTH.

One can interpret this finding such that below average contributors respond to the feedback inflation (after they are informed on the true nature of the manipulation) by lowering cooperation since they cannot foster the cooperation of above average contributors with their own contributions. In this sense, the publicly known feedback manipulation in LIE undermines cooperation in the entire group.

To gain more insight into the dynamics of contributions in consecutive periods, we analyze individual responses to received feedback in a Tobit regression model with individual random effects as model I. We analyze the dependent variable g_i^t which is censored within the limits 0 and 20 with periods $t \in \{22, 23, ..., 40\}$. We estimate the coefficients for the contribution in $t-1$, g_i^{t-1}, the dummy variables *lie* and *truth*, which are 1 if the observation comes from the LIE (TRUTH) treatment, and 0 otherwise.

In addition, we introduce two independent variables reflecting the received feedback. $feedback_L^{t-1}$ stands for feedback concerning the average contributions of the other group members in $t-1$ if i contributed *less* than the average in that round. That is, the coefficient for $feedback_L^{t-1}$ measures the response to the correct feedback across all treatments in terms of i's consecutive contribution. $feedback_A^{t-1}$ then is a dummy variable which is 1 if i contributes *above* average in $t-1$, and 0 otherwise. Thus, $feedback_A^{t-1}$ indicates the response of i who contributes above average, whereas the interaction terms $feedback_A^{t-1} \times lie$ and $feedback_A^{t-1} \times truth$ indicate deviations from the former response due to potentially manipulated feedback in LIE, and due to the inconvenient truth that other group members earned money at the expense of i in TRUTH.

Finally, we consider a linear time trend t, as cooperation rates typically decline towards the end of the experiment. As a robustness check, we estimate a level two mixed-effect maximum likelihood estimation with group fixed effects and individual random effects on the same set of independent variables in model II [14]. Estimation results are summarized in Table 1. We report estimated coefficients along with standard errors, number of observations (*obs*), and number of independent observations (*nobs*). Asterisks indicate significance levels. The fitness of the models are tested using Wald-χ^2-tests.

Table 1. Estimation results for a Tobit regression with individual random effects (model I) and a level two mixed-effect maximum likelihood estimation with group fixed effects and individual random effects (model II) with dependent variable g_i^t; coefficients are reported along standard errors in parenthesis; *** indicates significance at a $p < 0.01$ level, ** at a $p < 0.05$ level and * at a $p < 0.1$ level.

	Dependent Variable: g_i^t			
	Model I		**Model II**	
g_i^{t-1}	0.493 ***	(0.027)	0.422 ***	(0.019)
lie	−3.153 ***	(1.138)	−1.734 *	(0.908)
$truth$	−2.104 *	(1.117)	−0.890	(0.893)
$feedback_L^{t-1}$	0.141 ***	(0.034)	0.077 ***	(0.024)
$feedback_A^{t-1}$	0.689	(0.649)	0.271	(0.476)
$feedback_A^{t-1} \times lie$	−1.572 ***	(0.562)	−1.321 ***	(0.416)
$feedback_A^{t-1} \times truth$	−1.098 **	(0.542)	−1.327 ***	(0.402)
t	−0.114 ***	(0.015)	−0.074 ***	(0.012)
constant	8.771 ***	(1.088)	7.831 ***	(0.816)
obs	3724		3724	
nobs	196		49	
Wald-χ^2-test	620.14 ***		822.32 ***	

[14] Model II controls for dependency of contributions within a group, but cannot take into account that we have censored data in the dependent variable, while the opposite is true for model I.

When interpreting the results, we follow a conservative approach referring to the minimum of the significance level of model I and model II for estimated coefficients. We find strong evidence for a significant positive path dependency of contributions (g_i^{t-1}) while there is an overall negative time trend (t) in the data. Both findings are not surprising and in line with previous studies (e.g., Zelmer [17]). In addition, there is weak evidence that the overall contribution level in LIE is lower than in the other two treatments—despite the fact that contributions in the initial period of the second sequence (i.e., period 21) are not considered in the estimations.

Regarding the effect of received feedback, we find strong evidence for an adjustment of below average contributors: the higher the contributions of the other group members are on average, the more do contributions increase in round t (see the coefficient for $feedback_L^{t-1}$). This observation suggests conditional cooperation of below average contributors. With respect to above and at average contributors, there is a significant influence of the revelation of the feedback mechanism on the adjustment of contributions: while we find no significant coefficient for $feedback_A^{t-1}$, coefficients for the interaction term $feedback_A^{t-1} \times lie$ is significantly negative. Interestingly, observing the inconvenient truth in TRUTH leads to a similar response: χ^2-tests cannot reject the hypothesis that the coefficients $feedback_A^{t-1} \times lie$ and $feedback_A^{t-1} \times truth$ are the same ($\chi^2(1) = 1.09$, $p = 0.295$ for model I and $\chi^2(1) = 0.00$, $p = 0.985$ for model II).

Thus, both potentially inflated feedback and the inconvenient truth yield approximately the same response in terms of contributions. However, it could be that different motives lead to the decrease in contributions in TRUTH and in LIE; while above-average contributors in the former treatment reduce their contributions in order to avoid being a sucker, in the latter treatment it could be the mere uncertainty about size of contributions which drives participants to contribute more cautiously. In summary, combining the results from the earlier analysis on the difference between the initial contribution of the second and the first sequence (i.e., in period 21 and in period 1) with the regression results leads to the following:

Result 2. *Initial contributions in the second sequence of* LIE *are significantly lower than in both other treatments. However, above average contributors who receive potentially inflated feedback in* LIE *contribute similarly as above average contributors who receive truthful feedback in* TRUTH.

4. Conclusions

Cooperation achieved through feedback inflation cannot be sustained after the manipulation has been revealed; the development of contribution rates in the aftermath of the noble lie is independent of the treatment condition. Still being lied to basically brings out the same negative development of contributions as being told the truth after the lie is revealed: above average contributors decrease their contributions. Thus, in a sense, what goes around, comes around: contributions decline in the course of the second sequence for both treatment conditions.

However, there is an important difference between the two conditions: LIE yields significantly lower initial contributions than TRUTH (and BASELINE). In other words, after the manipulation is exposed, the prospect of playing another sequence of the VCM with inflated feedback leads to less cooperation right from the start. Importantly, this effect is driven not only by above average contributors; all players decrease their contributions, while contributions converge in period 21 in TRUTH (compare Figure 2). As a consequence, in LIE, this effect causes lower contributions throughout almost the entire second sequence of the game.

Overall, the lie is only "useful" before it has been revealed. Once exposed—and in an open society this is very likely to happen—it destroys cooperation. On the other hand, telling the inconvenient truth in the aftermath of a noble lie leads to similar responses as potentially inflated feedback. Thus, once a noble lie is uncovered, the truth prevails.

Acknowledgments: We thank Thomas Lauer and Bettina Rockenbach for providing detailed information on their experiment and data from some of their sessions. Financial support by the German Science Foundation (DFG) for Andreas Nicklisch (Ni 1610/1-1) and Kai-Uwe Schnapp (Schn 1282/4-1) is gratefully acknowledged.

Author Contributions: All authors contribued equally to all parts of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Instructions

We report the English translations of the instructions for the TRUTH treatment. Differences in the other treatments are marked by footnotes.

A.1. Instructions for Sequence 1

Welcome to the experiment!

You will take part in an experiment on economic decision-making. Please note that from now on and throughout the experiment **all forms of communication are strictly prohibited**. If you have a question, please indicate this by sticking your hand out of your cabin. All the decisions are taken anonymously, i.e., no other participant will know your identity. The payments at the end of the experiment will also be made anonymously, which means that no other participant will know how much your payment is.

Instructions for the experiment and general information

The experiment will consist of **two sequences**. First, you will only receive the instructions for sequence 1. The instructions for sequence 2 will be distributed after sequence 1 has been completed. Your decisions in sequence 1 will neither have an impact on your possible choices nor on your earnings in sequence 2.

Information about sequence 1

You are in a group consisting of four members in total. During all of sequence 1, you will only interact with members of your group.

Sequence 1

Contributions of group members

- In **sequence 1**, you will play 20 rounds and every round has the same structure.
- Every player receives an **endowment of 20 points** in each round.
- Every player has to decide how many of these 20 points he or she will **contribute** to the group project.
- All points contributed to the group project will be **multiplied by 1.6** and **equally** distributed amongst all four players, i.e., for every point that one player contributes to the group project, **every player gets 0.4** $(= 1.6/4)$ **points**.
- Points not contributed to the group project are kept by the individual player.

Information about the contributions of other group members

- From the second round on, you will be informed about the **average contribution** that your group members made in the previous round. The average will be rounded to the next **integer number**, so that you will only see feedback in whole numbers.

Please note: Apart from rounding, this information may <u>deviate from the actual value</u>.

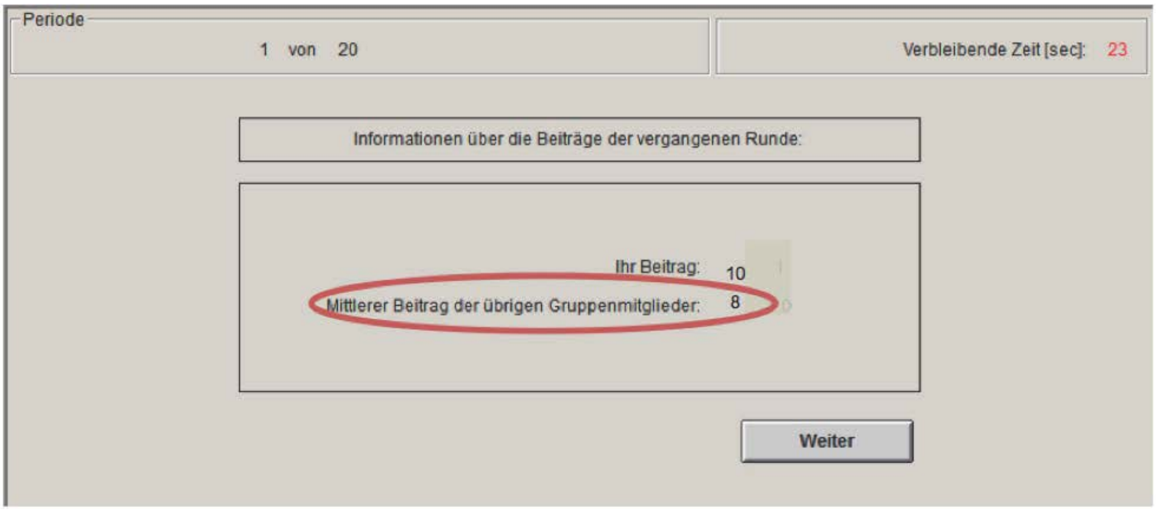

Figure A1. Screenshot–contribution phase.

Calculation of income per round

- Your income in each round will consist of two parts:
  - points which you did not contribute
  - your share of the group project
- **Please note:** For the calculation of your income, the actual contributions of your group members and not the announced values will be relevant.

Income per round:

$$\begin{aligned}&(\text{endowment } (20) - \text{your contribution to the group project})\\&+ \text{sum of } \textbf{all} \text{ contributions of your group} \times 1.6/4.\end{aligned}$$

Calculation example for the case that **you contributed 10**, while **your group members contributed 12, 8 and 4** points:

$$\begin{aligned}&= (20-10) + (10+12+8+4) \times 1.6/4\\&= 10 + 13.6\\&= 23.6.\end{aligned}$$

Income in sequence 1

At the end of sequence 1, we will randomly pick **one of the twenty rounds** to calculate your income. This means that every one of the twenty rounds played is possibly relevant to your payment. Your income from sequence 1 then is the points earned in the round randomly chosen, which will be paid according to an exchange rate of 1 Euro for every eight points.

Total income from the experiment

Your total income from the experiment will consist of the guaranteed 5 Euro, plus your income from sequence 1, plus your income from sequence 2.

Good luck!

A.2. Instructions for Sequence 2

Information about sequence 2:

Sequence 2 will consist of 20 more rounds of the game played in sequence 1. The composition of your group will not change.

Please note:

In sequence 1, if you contributed **less than the average contribution** of your group in the previous round, you were shown **the actual average contribution** of your group members.

In sequence 1, if you contributed **exactly the average contribution** of your group in the previous round, you were shown **the actual average contribution** of your group members.

In sequence 1, if you contributed **more than the average contribution** of your group in the previous round, you were shown **your own contribution as average contribution** of your group members.

However, in sequence 2, you will **always be shown the actual average contribution** of your group members [15].

Income in sequence 2

At the end of sequence 2, we will randomly pick **one of the second twenty rounds** to calculate your income. This means that every one of the twenty rounds played is possibly relevant to your payment. Your income from sequence 2 is then the points earned in the round randomly chosen, which will be paid according to an exchange rate of 1 Euro for every eight points.

Total income from the experiment

Your total income from the experiment will consist of the guaranteed 5 Euro, plus your income from sequence 1, plus your income from sequence 2.

Good luck!

References

1. Dixon, P. Political Skills or Lying and Manipulation? The Choreography of the Northern Ireland Peace Process. *Political Stud.* **2002**, *50*, 725–741.
2. Aughey, A. The Art and Effect of Political Lying in Northern Ireland. *Ir. Political Stud.* **2002**, *17*, 1–16.
3. Gneezy, U.; Rockenbach, B.; Serra-Garcia, M. Measuring lying aversion. *J. Econ. Behav. Organ.* **2013**, *93*, 293–300.
4. Hey, J.D. Experimental economics and deception. A comment. *J. Econ. Psychol.* **1998**, *19*, 397–401.
5. Hertwig, R.; Ortmann, A. Deception in experiments: Revisiting the arguments in its defense. *Ethics Behav.* **2008**, *18*, 59–62.
6. Ortmann, A.; Hertwig, R. The costs of deception: Evidence from psychology. *Exp. Econ.* **2002**, *5*, 111–131.
7. Jamison, J.; Karlan, D.; Schechter, L. To deceive or not to deceive: The effect of deception on behavior in future laboratory experiments. *J. Econ. Behav. Organ.* **2008**, *68*, 477–488.
8. Hoffmann, M.; Lauer, T.; Rockenbach, B. The royal lie. *J. Econ. Behav. Organ.* **2013**, *70*, 305–313.
9. Marwell, G.; Ames, R.E. Economists free ride, does anyone else?: Experiments on the provision of public goods. *J. Public Econ.* **1981**, *15*, 295–310.
10. Weimann, J. Individual behaviour in a free riding experiment. *J. Public Econ.* **1994**, *54*, 185–200.
11. Gneezy, U. Deception: The role of consequences. *Am. Econ. Rev.* **2005**, *95*, 384–394.
12. Sanchez-Pages, S.; Vorsatz, M. An experimental study of truth-telling in a sender–receiver game. *Games Econ. Behav.* **2007**, *61*, 86–112.
13. Serra-Garcia, M.; van Damme, E.; Potters, J. Hiding an inconvenient truth: Lies and vagueness. *Games Econ. Behav.* **2011**, *73*, 244–261.
14. Serra-Garcia, M.; van Damme, E.; Potters, J. Lying about what you know or about what you do? *J. Eur. Econ. Assoc.* **2013**, *11*, 1204–1229.
15. Engel, C.; Kube, S.; Kurschilgen, M. *Can We Manage First Impressions in Cooperation Problems. An Experimental Study on Broken (and Fixed) Windows*; Discussion Paper No. 5; Max Planck Institute for Research on Collective Goods: Bonn, Germany, 2011.

[15] In LIE, it is stated that "It will be the same in sequence 2," thereafter, the previous paragraph is repeated replacing sequence 1 with sequence 2. In BASELINE, the sentences starting from "Please note:" until "However, in sequence 2 you will always be shown the actual average contribution of your group members" are missing.

16. Irlenbusch, B.; Rilke, R.M. *Public Good Examples—On the Role of Limited Feedback in Voluntary Contribution Games*; Working Paper No. 4; Cologne Graduate School in Management, Economics and Social Sciences: Köln, Germany, 2013.
17. Zelmer, J. Linear public goods experiments: A meta-analysis. Zelmer Linear public goods experiments. *Exp. Econ.* **2003**, *6*, 299–310.
18. Chaudhuri, A. Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Exp. Econ.* **2011**, *14*, 47–83.
19. Fischbacher, U.; Gächter, S. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *Am. Econ. Rev.* **2010**, *100*, 541–556.
20. Fischbacher, U. z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* **2007**, *10*, 171–178.
21. Bock, O.; Nicklisch, A.; Baetge, I. hroot: Hamburg registration and organization online tool. *Eur. Econ. Rev.* **2014**, *71*, 117–120.

© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).