*Corrigendum*

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

# Corrigendum: Truth, Lies, and Gossip

Psychological Science
1–3

DOI: 10.1177/0956797620944135
www.psychologicalscience.org/PS
SAGE

Original article: Peters, K., & Fonseca, M. A. (2020). Truth, lies, and gossip. *Psychological Science, 31*, 702–714. doi:10.1177/0956797620916708

In the original article, the authors reported that they categorized 150 lies in avoidance rounds as positive. These included 40 lies stating that (a) the investor had sent a positive number of tokens and (b) the agent had returned more than 0% but less than 33% of received tokens. However, the authors now believe that these 40 lies do not easily fit into their typology, as it is not clear that these lies are more positive (or, indeed, more negative; i.e., more or less likely to lead to investment) than a true statement that the agent was avoided. They therefore reran their analyses, omitting these 40 lies. This Corrigendum is updating the values associated with these analyses as well as some passages describing them, all of which occur in the Results section. The authors feel that, if anything, omitting these lies strengthens their pattern of findings.

In the Descriptive Statistics subsection (p. 704), the second paragraph, fifth sentence, is being changed as follows, and the sentence in parentheses is being added:

> In avoidance rounds, when no tokens were sent, participants who told positive lies ($n$ = 110) claimed that an average of 7.76 tokens had been sent and 13.39 had been returned, and participants who told negative lies ($n$ = 166) claimed that an average of 8.07 tokens had been sent and 0 had been returned. (An additional 40 avoidance-round lies claimed that more than 0 but less than one third of tokens were returned; because these lies had an ambiguous valence, they were excluded from analyses that relied on this typology.)

In the Form and Function of Gossipers' Lies subsection, the 11th and 12th paragraphs (p. 709) are being changed as follows:

> This analysis revealed that payoffs to trustworthy agents were significantly lower when their behavior was misrepresented than when it was either truthfully described—control: $\chi^2(1) = 54.03$, $p < .001$; competition: $\chi^2(1) = 37.89$, $p < .001$—or exaggerated—control: $\chi^2(1) = 25.95$, $p < .001$; competition: $\chi^2(1) = 10.63$, $p = .001$. Payoffs from truth and exaggeration did not differ—control: $\chi^2(1) = 1.72$, $p = .190$; competition: $\chi^2(1) = 0.90$, $p = .344$.
>
> The reverse pattern was evident for payoffs to untrustworthy agents. Here, payoffs were significantly higher when agents' behavior was misrepresented than when it was either truthfully described—control: $\chi^2(1) = 9.93$, $p = .002$; competition: $\chi^2(1) = 40.89$, $p < .001$—or exaggerated—$\chi^2(1) = 35.20$, $p < .001$; competition: $\chi^2(1) = 69.93$, $p < .001$. Payoffs were higher under truth than exaggeration—control: $\chi^2(1) = 25.36$, $p = .001$; competition: $\chi^2(1) = 14.46$, $p < .001$. Thus, exaggeration is as effective as truth at achieving positive reciprocity and more effective at achieving negative reciprocity.

Also in the Form and Function of Gossipers' Lies subsection, the beginning of the 13th paragraph (p. 709) is being updated as follows (note that the first and last sentences of this paragraph, reproduced here, are remaining the same):

> As a final step, we ran this same analysis for investor payoffs (Figs. 4b and 4d). When interacting with trustworthy agents, investors who acted on the truth received higher payoffs than those who acted on misrepresentation—control: $\chi^2(1) = 34.77$, $p < .001$; competition: $\chi^2(1) = 21.20$, $p < .001$. In the control condition, payoffs from exaggeration did not differ from payoffs from truth, $\chi^2(1) = 0.02$, $p = .898$, and were significantly higher than payoffs from misrepresentation, $\chi^2(1) = 10.67$, $p = .001$. In competition, payoffs from exaggeration were significantly lower than payoffs from truth, $\chi^2(1) = 7.07$, $p = .008$, and did not differ from

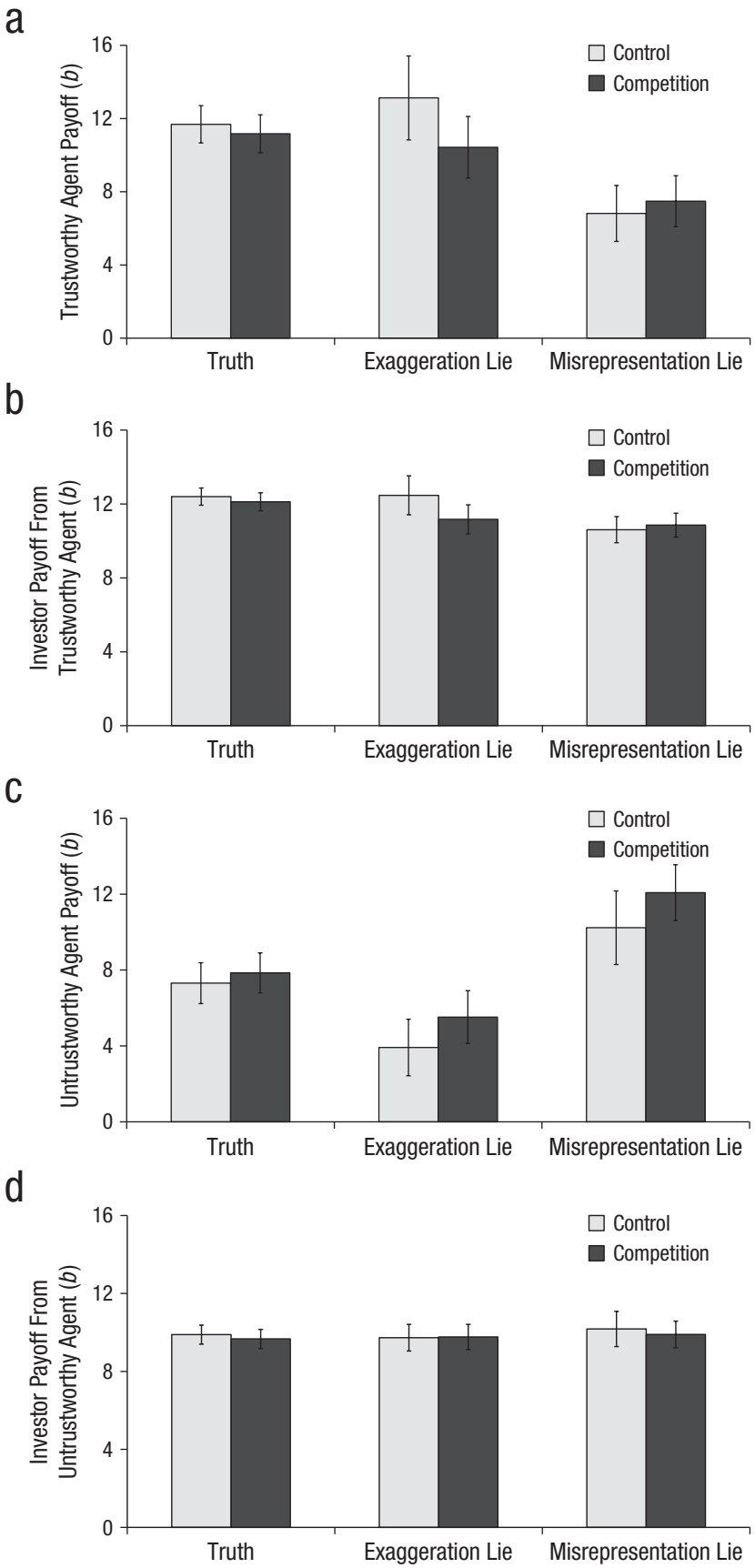


**Fig. 4.** Results of the regression on mean payoff as a function of agent's trustworthiness in the previous round, message type, and condition. Results are shown separately for analyses in which the dependent variable was trustworthy agents' payoffs (a), investors' payoffs from trustworthy agents (b), untrustworthy agents' payoffs (c), and investors' payoffs from untrustworthy agents (d). Error bars represent 95% confidence intervals.

payoffs from misrepresentation, $\chi^2(1) = 0.57$, $p = .450$. When investors interacted with untrustworthy agents, investor payoffs did not significantly differ on the basis of the content of the message in either the control condition, all $\chi^2$s$(1) \leq 0.85$, $p \geq .357$, or the competition condition, all $\chi^2$s$(1) \leq 0.57$, $p \geq .449$. Thus, truth (and, to a more limited extent, exaggeration) improves investor payoffs relative to misrepresentation when agents are trustworthy but does not when agents are untrustworthy.

Figures 4b and 4d are being replaced as well.

*Research Article*

Psychological Science
2020, Vol. 31(6) 702–714

DOI: 10.1177/0956797620916708
www.psychologicalscience.org/PS

# Truth, Lies, and Gossip

**Kim Peters[1,2] and Miguel A. Fonseca[1,3]**
[1]University of Exeter Business School; [2]School of Psychology, University of Queensland; and [3]Centre for Research in Economics and Management (NIPE), University of Minho

## Abstract

It is widely assumed that people will share inaccurate gossip for their own selfish purposes. This assumption, if true, presents a challenge to the growing body of work positing that gossip is a ready source of accurate reputational information and therefore is welfare improving. We tested this inaccuracy assumption by examining the frequency and form of spontaneous lies shared between gossiping members of networks playing a series of one-shot trust games ($N$ = 320). We manipulated whether gossipers were or were not competing with each other. We showed that lies make up a sizeable minority of messages and are twice as frequent under gossiper competition. However, this had no discernible effect on trust levels. We attribute this to the findings that (a) gossip targets are insensitive to lies and (b) some lies are welfare enhancing. These findings suggest that lies need not prevent—and may help—gossip to serve reputational functions.

## Keywords

gossip, accuracy, lies, trust, competition, reciprocity, open data, open materials, preregistered

Received 2/9/19; Revision accepted 1/27/20

> The man who comes with a tale about others has himself an axe to grind.
>
> —Chinese proverb

Within the body of cultural knowledge, there are ample warnings, such as the above proverb, about the dangers of attending to gossip. Such warnings are present in the academic literature, too. There it has been argued that people will share inaccurate gossip for their own selfish purposes, such as undermining enemies, promoting allies, or competing for mates (Hess & Hagen, 2006; Mace et al., 2018; McAndrew & Milenkovic, 2002). However, this argument, if true, presents a challenge to the growing body of work positing that as a ready source of accurate reputational information, gossip is able to bolster overall levels of cooperation (Dunbar, 1993).

Gossip is the class of communicated content that conveys information about the behaviors and characteristics of social actors (Peters & Kashima, 2015; Smith, 2014). Thus, it has the potential to inform us about our social world and the reputations of the people who inhabit it. If people act on the basis of the gossip they hear, cooperating more with individuals who are said to have behaved cooperatively in the past, then there are incentives for engaging in the costly cooperative acts that build positive reputation (Barclay, 2012; Nowak & Sigmund, 1998; Wedekind & Milinski, 2000). In other words, gossip may enable the indirect reciprocity that has been argued to boost cooperation in large social groups (e.g., neighborhoods, organizations, online communities, and markets; Alexander, 1987; Nowak & Sigmund, 2005; for a discussion of indirect-reciprocity theorizing and gossip, see Giardini & Wittek, 2019).

However, if gossip is inaccurate, so that information about a person's cooperation in the past is a poor guide to their cooperation in the future, then this virtuous cycle could break down (Giardini, 2012; Roberts, 2008; Smith, 2014). In line with this possibility, recent findings from a lab study (Fonseca & Peters, 2018; see also D. Fehr & Sutter, 2019) showed that gossip was less effective at securing cooperation when there was (and was known to be) a high chance that messages would be misdelivered and, hence, describe the wrong person's

**Corresponding Authors:**
Kim Peters, University of Exeter Business School, Rennes Dr., Exeter, EX4 4PU, United Kingdom
E-mail: k.o.peters2@exeter.edu.au

Miguel A. Fonseca, University of Exeter Business School, Rennes Dr., Exeter, EX4 4PU, United Kingdom
E-mail: m.a.fonseca@exeter.ac.uk

previous cooperation. Interestingly though, there was no evidence that gossip was less effective when this inaccuracy was spontaneously introduced by gossipers. Therefore, it is currently unclear whether inaccuracy—especially that which occurs naturally—is indeed one of the main threats to the capacity of gossip to fulfill reputational functions. To shed light on this, we need a better understanding of when and why gossipers lie.

In this article, we summarize the results of a preregistered behavioral study that was designed to build this understanding by examining the spontaneous lies shared between gossiping members of networks playing a series of one-shot trust games. We tested whether competition between gossipers increases lying, a prediction supported by game theory (Crawford & Sobel, 1982), and whether this increase would spark a low-discrimination, low-trustworthiness, and low-trust cascade. We tested the novel hypothesis that as gossiper–audience competition increases, gossip will become less accurate (Hypothesis 1). Further, replicating findings from a similar paradigm by Fonseca and Peters (2018), we also tested the hypothesis that as gossip becomes less accurate, there will be less gossip-based discrimination (Hypothesis 2a). This will lead to less trustworthiness (Hypothesis 2b) and in turn to less trust (Hypothesis 2c).

We also examined the form and functions of gossipers' lies. This exploratory analysis revealed, for the first time, that lies are used to pursue negative and positive social welfare goals and that some lies may serve reputational functions more effectively than truth.

## Method

### *Participants*

We recruited 320 participants (48% male; age: $M = 20.30$ years, $SD = 3.93$) from a pool of registered participants of the UQEEL lab at the University of Queensland and leaflets distributed on campus. We recruited the largest sample that was feasible in light of limits to participant availability and funding. The resulting sample included more than twice the number of participants per condition than previous work with this paradigm (Fonseca & Peters, 2018). Participants were paid in Australian dollars (AUD) an average of AUD$19.94 for their participation. This study received ethics approval from the University of Queensland. Preregistration information, materials, data, analysis code, and supplementary analyses are available at https://osf.io/k3jsk/.

### *Procedure*

The experiment was conducted in groups of 16 individuals who played 20 rounds of an anonymous trust game (Berg, Dickhaut, & McCabe, 1995) on networked computers. Half of the participants in each group were assigned to the role of investor, and the remainder were assigned to the role of agent. In each round, investors decided how much of their 10-token endowment to send to their allocated agent—a measure of trust. Agents received 3 times the number of tokens that were sent and decided how many to return to the investor—a measure of trustworthiness. Investors could, if they wanted to, avoid their agent by sending 0 tokens. Participants knew that they would never play against the same person twice in a row but that pairings were otherwise random.

At the end of each round, decisions and payoffs were presented on screen. To elicit gossip, we asked investors to send a message to the investor who would play with their agent in the next round. In this message, they stated the number of tokens that they had sent and the number that their agent had returned to them (the factual behavior remained on screen for reference). From the second round onward, investors received the message that described how their new agent had behaved in the previous round before deciding how many tokens to send. Agents knew that messages about their behavior were exchanged, but they never saw the messages (consistent with definitions of gossip as involving communications about absent third parties; Peters & Kashima, 2015; Smith, 2014).

To test whether competition could increase rates of lying, we allocated 10 groups of participants to the competition condition (the remaining 10 groups were in the control condition). To achieve this, we assigned all participants a color that they retained throughout the experiment (in each group, four investors and four agents were red, and four investors and four agents were blue). In the competition condition, investors competed along color lines for a bonus. Specifically, the total payoffs accumulated by red and blue investors over the course of the experiment were compared, and investors belonging to the winning color each received a bonus of AUD$5.00 (the losing color received AUD$0). Color had no payoff consequences for investors in the control condition or for agents in either condition (these participants received a flat bonus of AUD$2.50). Investors and agents received different information about the bonuses in their group (neither was informed about the existence of different conditions). In particular, investors were always told about the basis for all bonuses (e.g., in the control condition, they were told that investors and agents receive AUD$2.50), but agents were told only about their own bonus allocation (i.e., AUD$2.50) and that investor bonuses may be calculated differently. Agents, therefore, had no reason for anticipating different investor behavior in the competition and control conditions.

After completing the experiment, participants were paid (payment consisted of a show-up fee plus three

randomly selected rounds plus a bonus). They were then asked to complete a short survey. Among other things, participants reported their social bonding with same-color and different-color investors and agents (four sets of three items from the study by Peters & Kashima, 2007; $\alpha$ = .86–.92): "I had a social bond with [target]," "I connected with [target]," and "I trusted [target]." Investors also responded to three open questions asking when and why they had described an agent's behavior "truthfully," "too positively," and "too negatively." Agents instead responded to two items about their reputation concern ($r$ = .52, $p$ < .001): "When deciding how many tokens to return, I thought about what the next Investor would think of me," and "I returned more tokens than I wanted to in order to ensure that the next Investor would see me positively." They also responded to two sets of three items about their expectations of discrimination from same-color and different-color investors (collapsed into a six-item scale: $\alpha$ = .92; note that throughout this article, *discrimination* refers to the extent to which investors base their decision to trust an agent on the information they receive about this agent's previous trustworthiness): "I think that [color] investors decided how many tokens to send me on the basis of the message they received about my behaviour," "I think that if [color] investor was told that I returned a small number of tokens, they would send me fewer," and "I think that if [color] investor was told that I returned a large number of tokens, they would send me more." Items were rated on 7-point scales (1 = *strongly disagree*, 7 = *strongly agree*).

## Results

### Descriptive statistics

A slight majority of rounds involved a trusting investor, who sent a positive number of tokens (74% of rounds; tokens sent: $M$ = 6.97, $SD$ = 3.05), and a trustworthy agent, who returned at least one third of received tokens (71% of trust rounds; tokens returned: $M$ = 45.21%, $SD$ = 10.84). In the remaining rounds, agents were either untrustworthy (harming the investor by returning fewer than one third of received tokens; tokens returned: $M$ = 11.36%, $SD$ = 9.85) or were expected to be untrustworthy and thus avoided: Avoidance rounds were usually preceded by a message stating that the agent had been avoided (23% of cases) or that the agent had been untrustworthy (49% of cases; tokens returned: $M$ = 19.00%, $SD$ = 28.81).

Any message that misstated tokens sent or returned was considered to be a lie. Although investors told the truth most of the time, as was observed by Fonseca and Peters (2018), a substantial minority of messages were lies (25.88%). These lies were substantial in size. In trust rounds, participants who told positive lies ($n$ = 171) claimed that an average of 7.93 tokens had been sent and 14.56 returned, when the actual values were 6.84 sent and 7.44 returned; participants who told negative lies ($n$ = 341) claimed that an average of 7.87 tokens had been sent and 3.38 returned, when the actual values were 7.22 sent and 8.68 returned. In avoidance rounds, when no tokens were sent, participants who told positive lies ($n$ = 110) claimed that an average of 7.76 tokens had been sent and 13.39 had been returned, and participants who told negative lies ($n$ = 166) claimed that an average of 8.07 tokens had been sent and 0 had been returned. (An additional 40 avoidance-round lies claimed that more than 0 but less than one third of tokens were returned; because these lies had an ambiguous valence, they were excluded from analyses that relied on this typology.) For more details of these analyses and all of those that follow, see the Supplemental Material available online.

### Manipulation check

To test the effectiveness of our competition manipulation, we ran a 2 (condition: competition, control) × 2 (target color: same, different) mixed analysis of variance (ANOVA) of investors' reported social bonds. The main effects of target color, $F(1, 157) = 141.75$, $p < .001$, $\eta^2 = .35$, and condition, $F(1, 157) = 14.08$, $p < .001$, $\eta^2 = .08$, were qualified by the two-way interaction, $F(1, 157) = 69.74$, $p < .001$, $\eta^2 = .20$. This revealed that the tendency for investors to report a stronger social bond with in-group (same color) than out-group (different color) investors was more pronounced in the competition condition (in-group: $M$ = 5.29, $SD$ = 1.60; out-group: $M$ = 2.94, $SD$ = 1.30) than in the control condition (in-group: $M$ = 3.60, $SD$ = 1.43; out-group: $M$ = 3.28, $SD$ = 1.45). This suggests that the manipulation introduced a stronger intergroup dynamic in the competition condition than in the control condition.

### Hypothesis testing

We first report the tests of our hypotheses that participants would be especially likely to lie if they were competing with their audience and that this would lead to a breakdown in discrimination, trust, and trustworthiness (preregistered analyses produced identical findings and are reported in the Supplemental Material). To understand how competition affected rates of lying, we ran a mixed-effects regression of whether or not the message was a lie on a condition dummy, an audience-group-affiliation dummy, and their interaction. In line with Hypothesis 1, lies were most prevalent when participants communicated with out-groups in competition. Specifically, in the competition condition, 43% of

the messages that investors sent to out-group audiences were lies, whereas only 20% of those they sent to in-group audiences were lies, $\chi^2(1) = 192.19$, $p < .001$. In the control condition, in contrast, investors were about as likely to lie to the in-group as to the out-group (21% and 19% of messages, respectively), $\chi^2(1) = 1.03$, $p = .310$. It is interesting to note that even in the absence of incentives, one out of every five messages was a lie—an almost identical number to that observed by Fonseca and Peters (2018).

To determine whether investors were less likely to discriminate on messages when lies were more prevalent, we ran a mixed-effects regression of investors' trust on a condition dummy, a gossiper-group-affiliation dummy, the content of the message (expressed as proportion of tokens returned), and all two- and three-way interactions. In line with Hypothesis 2a, we found a significant three-way interaction among condition, gossiper's group affiliation, and message content; investors' tendencies to act on gossip were more influenced by the gossiper's group affiliation in the competition condition than in the control condition, $\chi^2(1) = 5.18$, $p = .023$. Specifically, in the competition condition, a message that an agent had returned half of received tokens (vs. none) increased trust by an average of 3.39 tokens if it came from an in-group gossiper but only by 1.71 tokens if it came from an out-group gossiper, $\chi^2(1) = 22.94$, $p < .001$. In the control condition, there was no evidence that investors responded differently to messages from in-group or out-group gossipers, $\chi^2(1) = 0.79$, $p = .375$, and the above message increased trust by 5.17 tokens on average.

This suggests that the association between agents' behavior in one round and the number of tokens they received in the next was attenuated in the competition condition. To determine whether agents took advantage of this opportunity to behave selfishly, we ran a mixed-effects regression of the proportion of received tokens that agents returned to investors on a condition dummy. Unexpectedly, and contrary to Hypothesis 2b, results showed that agents were about as trustworthy in lower-discrimination competition networks (return: $M = 34\%$) as they were in higher-discrimination control networks (return: $M = 37\%$), $\chi^2(1) = 1.50$, $p = .220$. This behavior corresponded with our finding that agents' self-reported concern for their reputation (control: $M = 5.51$, $SD = 1.44$; competition: $M = 5.31$, $SD = 1.47$) and expectations of discrimination (control: $M = 4.89$, $SD = 1.41$; competition: $M = 4.66$, $SD = 1.52$) were reasonably high and did not vary significantly with condition, all $t$s$(156) \leq 0.98$, $p \geq .327$. In other words, the actual discrimination that agents experienced did not appear to alter their beliefs about this discrimination. In light of our finding that competition agents were no less trustworthy than control agents, it is unsurprising to find that levels of trust did not appear to differ across condition either, contrary to Hypothesis 2c, $\chi^2(1) = 0.21$, $p = .647$.

These findings are consistent with claims that lies are an important component of gossip, especially if there are material incentives for lying. However, whereas lies were 63% more prevalent in the competition condition (and levels of discrimination were commensurably lower), trustworthiness and trust were not significantly affected, which suggests that gossip targets are not especially adept at calibrating their levels of reputation concern to gossipers' tendencies to act on the gossip they receive. The nontrivial base rate of lies also suggests that a desire to mislead the audience is not the only reason that gossipers lie. Indeed, Fonseca and Peters (2018) noted that some gossipers provided unprompted descriptions of using lies for a range of purposes, including punishing or rewarding targets and boosting investment. To shed light on the range of purposes that lies may serve, we now systematically explore their form and function.

## *Form and function of gossipers' lies*

We identified four main forms of lies. These reflected the interaction of two orthogonal dimensions: first, whether the agent in question was trustworthy or untrustworthy (i.e., returned at least 33% of received tokens or returned less than this), and second, whether the lie claimed that the agent was more or less trustworthy than the agent actually had been. *Positive-misrepresentation lies* described untrustworthy or avoided agents as more trustworthy than they actually were, *negative-misrepresentation lies* described trustworthy agents as less trustworthy than they actually were, *positive-exaggeration lies* described trustworthy agents as more trustworthy than they actually were, and *negative-exaggeration lies* described untrustworthy or avoided agents as less trustworthy than they actually were.

As a first step toward understanding why gossipers chose to share these different kinds of lies, we independently coded investors' postexperimental explanations for their decisions to send accurate or inaccurate messages. This analysis revealed that between 80% and 98% of codable explanations related to social welfare motives—that is, a desire to send content to help or harm the agent or audience (coder $\kappa$s = .65–.94). Social welfare motives representing more than 5% of explanations for a given type of message are summarized in Figure 1.

This analysis points to an important distinction between misrepresentation lies and exaggeration lies. The former were solely justified by a desire to harm the audience (by encouraging behavior likely to diminish the audience's payoffs). In contrast, the latter were

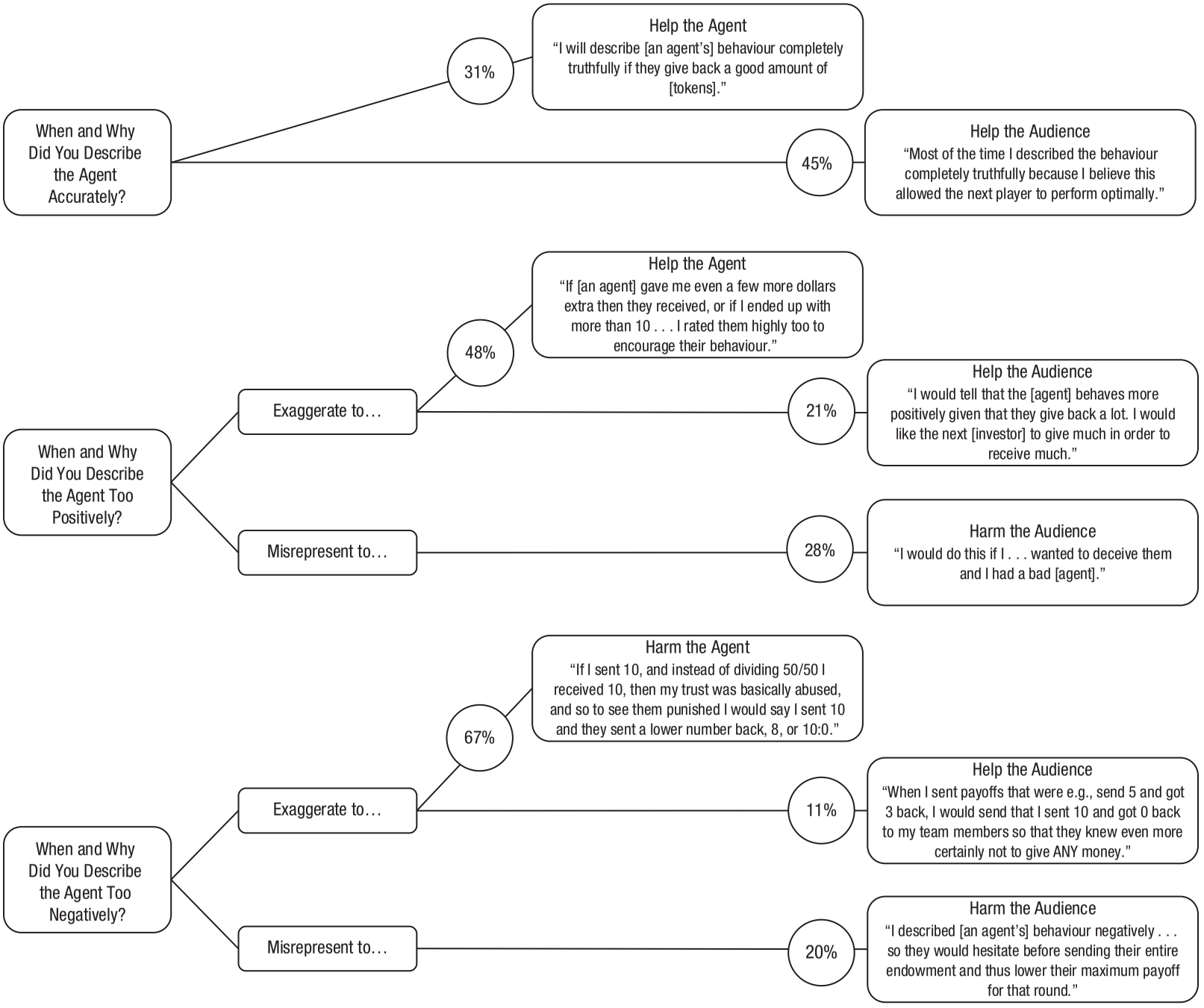


**Fig. 1.** Percentage of participants who used social welfare themes when explaining when and why they used different gossip content. For each question, motives about the welfare of the agent and motives about the welfare of the audience are shown separately. Typical responses are provided. Percentages were calculated from investors who reported sending a given type of message and provided a codable explanation—accurate: $n = 122$, positive inaccurate: $n = 67$, and negative inaccurate: $n = 79$.

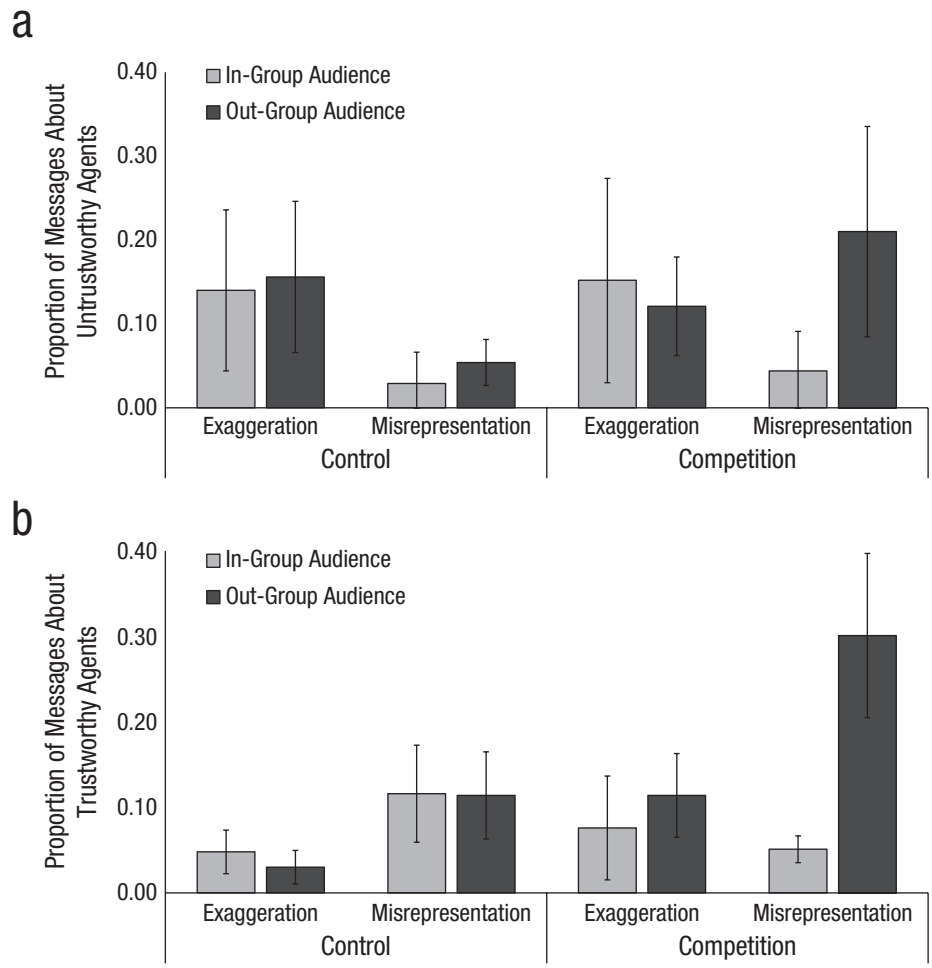


**Fig. 2.** Estimated proportion of trust-round lies as a function of condition, lie type, and audience, separately for messages about (a) untrustworthy agents and (b) trustworthy agents. Error bars represent 95% confidence intervals.

justified by a desire to help the audience (by encouraging behavior likely to improve the audience's payoffs) or to achieve reciprocity with the agent (by encouraging behavior likely to improve the payoffs of trustworthy agents and diminish those of untrustworthy agents).

This suggests that the increased prevalence of lies in the competition condition should be primarily underpinned by an increase in misrepresentation (associated with the desire to harm competitor audiences). It also suggests that the lies that were sent to other (noncompetitor) audiences should be primarily composed of exaggeration. To test this behavioral expectation, we used a multinomial logit regression of the type of lie on agent type (trustworthy, untrustworthy), a condition dummy, a gossiper-group-affiliation dummy, and all two-way and three-way interactions. We clustered standard errors at the session level to account for interdependencies (Wooldridge, 2003). Figure 2 displays the results of the estimation.

The analysis presented here relates only to trust rounds, but the pattern in avoidance rounds was generally consistent (see the Supplemental Material). In line with the possibility that misrepresentation lies were motivated by a desire to harm the audience, we found that gossipers were most likely to tell misrepresentation lies to out-group competitors. Specifically, we found that competition investors made significantly more use of misrepresentation when messaging the out-group than control investors did—untrustworthy agents: $\chi^2(1) = 5.36$, $p = .021$; trustworthy agents: $\chi^2(1) = 11.56$, $p < .001$. Competition investors also made significantly more use of misrepresentation lies when messaging the out-group than the in-group—untrustworthy agents: $\chi^2(1) = 11.10$, $p < .001$; trustworthy agents: $\chi^2(1) = 27.82$, $p < .001$.

To examine the kinds of lies that were sent to noncompetitor audiences, we first looked at in-group audiences in the competition condition (with whom

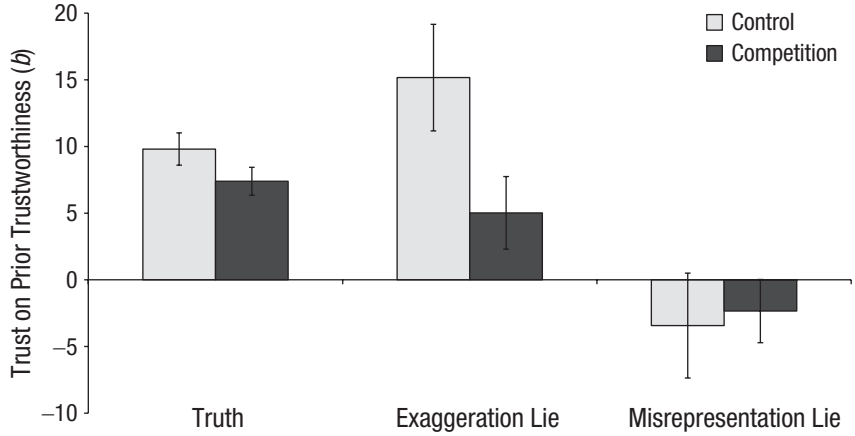


**Fig. 3.** Results of the regression on reciprocation (the degree to which an agent's trustworthiness in one round was reciprocated in the next) as a function of message type and condition. Error bars represent 95% confidence intervals.

investors reported having a strong social bond). Lies to this audience were never significantly more frequent than to out-group competitors (see the tests for misrepresentation in the preceding paragraph)—exaggerating trustworthiness: $\chi^2(1) = 2.47$, $p = .116$; exaggerating untrustworthiness: $\chi^2(1) = 0.32$, $p = .574$. However, we did find that when competition investors lied to the in-group, they were more likely to exaggerate than misrepresent—although this difference was significant only when communicating about untrustworthy agents, $\chi^2(1) = 5.61$, $p = .018$; trustworthy agents: $\chi^2(1) = 0.78$, $p = .376$. This is consistent with the possibility that investors may have used exaggeration lies in an attempt to help the in-group audience.

Next, we looked at the lies that investors told in the control condition, which revealed a different pattern again. Here, we found that investors showed a preference for telling negative rather than positive lies (e.g., exaggerating untrustworthiness and misrepresenting trustworthiness) regardless of the identity of the audience—untrustworthy target: in-group audience, $\chi^2(1) = 4.80$, $p = .029$, out-group audience, $\chi^2(1) = 3.84$, $p = .050$; trustworthy target: in-group audience, $\chi^2(1) = 3.78$, $p = .052$, out-group audience, $\chi^2(1) = 14.04$, $p < .001$. This negativity bias was robust across more stringent standards for trustworthiness (i.e., where an agent had to return 40%, 45%, or 50% of tokens to be considered trustworthy), suggesting that it is not due to participants having higher standards for trustworthiness than we do (i.e., a 33% return rate).

Our analysis suggests that gossipers believe that misrepresentation impedes adaptive gossip-based discrimination but that exaggeration facilitates it by helping audiences to reciprocate target behavior. If true, this raises a novel possibility: that some lies may actually support the indirect reciprocity that has been implicated in cooperation in large populations. To test this possibility, we analyzed how the degree to which an agent's trustworthiness in one round was reciprocated in the next was affected by whether the message about the agent's behavior was the truth, an exaggeration lie, or a misrepresentation lie. We ran a mixed-effects regression of the tokens sent to an agent on a condition dummy, two lie dummies (exaggeration or misrepresentation; truth was the omitted category), the agent's trustworthiness in the previous round (i.e., proportion of tokens returned), and the two- and three-way interactions among condition, each lie dummy, and trustworthiness. The reciprocation coefficients are graphed in Figure 3.

Starting with the control condition, we found that the association between agents' trustworthiness and the extent to which they were subsequently trusted was positive and significant when investors told the truth, $\chi^2(1) = 253.50$, $p < .001$. Importantly, when investors exaggerated the agent's trustworthiness, this association was significantly stronger, $\chi^2(1) = 55.36$, $p < .001$. In contrast, when investors misrepresented the agent's trustworthiness, the association was weakly negative, $\chi^2(1) = 2.91$, $p = .088$. Turning to the competition condition, we found that the association between agents' trustworthiness and the extent to which they were subsequently trusted was positive and significant when investors told the truth, $\chi^2(1) = 191.33$, $p < .001$, or exaggerated, $\chi^2(1) = 13.15$, $p < .001$; these did not significantly differ, $\chi^2(1) = 2.53$, $p = .111$. When investors misrepresented the agent's trustworthiness, the association was again weakly negative, $\chi^2(1) = 3.72$, $p = .054$.

In short, these results suggest that truthful gossip and exaggeration lies are both likely to ensure that agents receive their just desserts, whereas misrepresentation lies are likely to prevent this. To test this

possibility, we ran a mixed-effects regression of agent payoffs on a condition dummy, two lie dummies (exaggeration or misrepresentation), an agent-trustworthy-type dummy, and the two- and three-way interactions among condition, each lie dummy, and agent type. The coefficients are graphed in Figures 4a and 4c.

This analysis revealed that payoffs to trustworthy agents were significantly lower when their behavior was misrepresented than when it was either truthfully described—control: $\chi^2(1) = 54.03$, $p < .001$; competition: $\chi^2(1) = 37.89$, $p < .001$—or exaggerated—control: $\chi^2(1) = 25.95$, $p < .001$; competition: $\chi^2(1) = 10.63$, $p = .001$. Payoffs from truth and exaggeration did not differ—control: $\chi^2(1) = 1.72$, $p = .190$; competition: $\chi^2(1) = 0.90$, $p = .344$.

The reverse pattern was evident for payoffs to untrustworthy agents. Here, payoffs were significantly higher when agents' behavior was misrepresented than when it was either truthfully described—control: $\chi^2(1) = 9.93$, $p = .002$; competition: $\chi^2(1) = 40.89$, $p < .001$—or exaggerated—$\chi^2(1) = 32.50$, $p < .001$; competition: $\chi^2(1) = 69.93$, $p < .001$. Payoffs were higher under truth than exaggeration—control: $\chi^2(1) = 25.36$, $p = .001$; competition: $\chi^2(1) = 14.46$, $p < .001$. Thus, exaggeration is as effective as truth at achieving positive reciprocity and more effective at achieving negative reciprocity.

As a final step, we ran this same analysis for investor payoffs (Figs. 4b and 4d). When interacting with trustworthy agents, investors who acted on the truth received higher payoffs than those who acted on misrepresentation—control: $\chi^2(1) = 34.77$, $p < .001$; competition: $\chi^2(1) = 21.20$, $p < .001$. In the control condition, payoffs from exaggeration did not differ from payoffs from truth, $\chi^2(1) = 0.02$, $p = .898$, and were significantly higher than payoffs from misrepresentation, $\chi^2(1) = 10.67$, $p = .001$. In competition, payoffs from exaggeration were significantly lower than payoffs from truth, $\chi^2(1) = 7.07$, $p = .008$, and did not differ from payoffs from misrepresentation, $\chi^2(1) = 0.57$, $p = .450$. When investors interacted with untrustworthy agents, investor payoffs did not significantly differ on the basis of the content of the message in either the control condition, all $\chi^2$s$(1) \leq 0.85$, $p \geq .357$, or the competition condition, all $\chi^2$s$(1) \leq 0.57$, $p \geq .449$. Thus, truth (and, to a more limited extent, exaggeration) improves investor payoffs relative to misrepresentation when agents are trustworthy but does not when agents are untrustworthy.

These exploratory findings suggest that gossipers use lies to achieve a range of outcomes and challenge the widespread assumption that lies are necessarily malicious and harmful. Next, we describe the results of a study in which we aimed to independently verify gossipers' motives for telling the truth, exaggeration lies, or misrepresentation lies.

## *Verifying the motives underlying gossipers' lies*

We used an online experiment ($N = 81$ UK-based adults; for full details of the method and results, see the Supplemental Material) that emulated the main study with a strategy method to independently verify the mapping of lies to social welfare motives. Participants responded to an identical set of questions three times. In the first neutral-audience iteration, participants were asked to imagine being an investor in a population that played repeated trust games with agents and exchanged messages about these interactions. Participants were then asked to imagine interacting with three agents in turn: an untrustworthy agent (8 tokens sent, 3 tokens returned), a trustworthy agent (8 tokens sent, 10 tokens returned), and an agent they chose to avoid (0 tokens sent). In each case, they were asked to rate their social welfare motives toward (a) their agent and (b) their audience (i.e., the next investor to play their agent; 7-point Likert-type scales: 1 = *strong desire to harm*, 7 = *strong desire to help*). They were also presented with three concrete messages that they could send to their audience (the truth, a positive lie, and a negative lie) and asked to rate how much each message would allow them to jointly affect their agent and audience as desired (8-point Likert-type scales: 0 = *not at all*, 7 = *definitely*). The content of the lies was based on the typical content sent about the three types of agents in the main study, rounded to the nearest feasible integer.

This was followed by two intergroup iterations. Specifically, participants were informed that they had joined a new population that consisted of two teams of investors who were competing to earn the most tokens. They were asked to respond to the above questions twice: once for noncompetitor audiences (i.e., in-group members) and once for competitor audiences (i.e., out-group members). The order of audience was randomized.

To check that the agent and audience manipulations affected participants' social welfare motives in the expected ways, we first regressed participants' desire to help (vs. harm) their agent on two dummies representing agent trustworthiness (trustworthy or untrustworthy; avoided was the omitted category), clustering standard errors at the level of the participant. This analysis revealed that relative to avoided agents, participants were less motivated to help untrustworthy agents ($b = -0.71$, 95% confidence interval, or CI = [−0.94, −0.47], $p < .001$) and more motivated to help trustworthy agents ($b = 0.59$, 95% CI = [0.34, 0.83], $p < .001$). We next regressed participants' desire to help (vs. harm) their audience onto two dummies representing

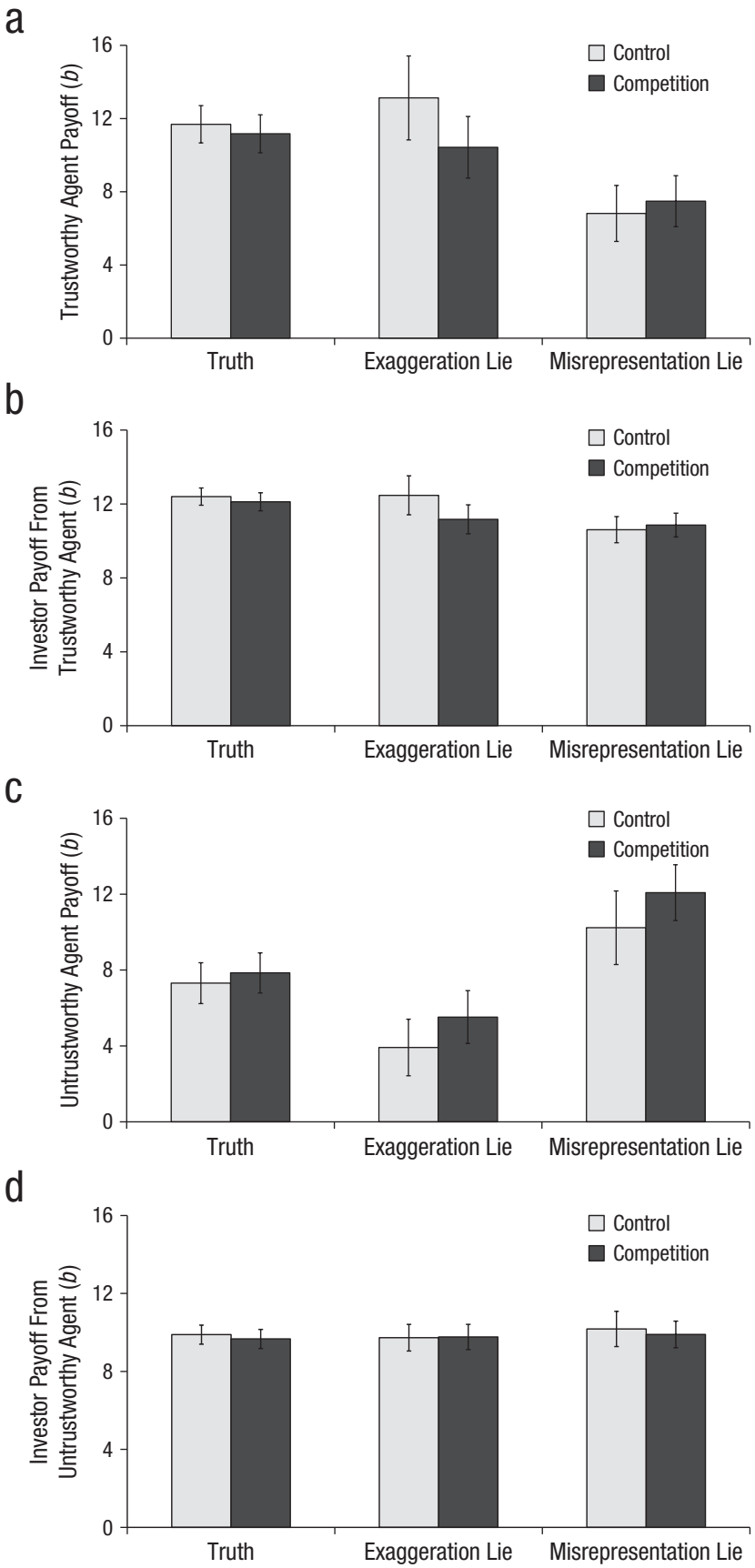


**Fig. 4.** Results of the regression on mean payoff as a function of agent's trustworthiness in the previous round, message type, and condition. Results are shown separately for analyses in which the dependent variable was trustworthy agents' payoffs (a), investors' payoffs from trustworthy agents (b), untrustworthy agents' payoffs (c), and investors' payoffs from untrustworthy agents (d). Error bars represent 95% confidence intervals.

audience affiliation (in-group or out-group; neutral was the omitted category), clustering standard errors at the participant level. This revealed that relative to neutral audiences, participants were less motivated to help out-group audiences ($b = -2.16$, 95% CI = [−2.53, −1.80], $p < .001$) and more motivated to help in-group audiences ($b = 0.67$, 95% CI = [0.42, 0.91], $p < .001$). The manipulations were therefore successful.

To understand participants' mapping of social welfare motives onto different messages, we ran regressions of participants' ratings that a given message would achieve their social motives for truthful messages and positive and negative lies in turn. To simplify interpretation, we ran the regressions once accounting for agent type and a second time accounting for audience affiliation. In each case, we regressed message-motive achievement onto participants' desire to help (vs. harm) their agent and help (vs. harm) their audience, two dummies representing either agent trustworthiness or audience affiliation, and the four two-way interactions between social welfare motives and the relevant dummies. Standard errors were clustered at the participant level.

The coefficients from these regressions are graphed in Figure 5. We first describe participants' perceptions that different messages can achieve their goals for the agent's social welfare (Figs. 5a and 5c). The analysis revealed that the truth was seen to help trustworthy agents ($p < .001$; all other coefficients, $p \geq .133$). Positive lies were seen to help agents when the audience was neutral ($p = .001$; all other coefficients, $p \geq .074$). Negative lies were seen to harm agents when they exaggerated agents' untrustworthiness ($p = .012$) or the audience was either neutral or an in-group member (all $p$s $< .001$). They were seen to help agents when the audience was an out-group member ($p = .035$; all other coefficients, $p \geq .126$).

We turned next to participants' perceptions that different messages can achieve their goals for the audience's social welfare (Figs. 5b and 5d). The analysis revealed that the truth was seen to help the audience regardless of its affiliation or the agent's trustworthiness (all $p$s $< .001$). Positive lies were seen to harm the audience when they misrepresent an avoided or untrustworthy agent ($p < .001$) or the audience was an in-group member ($p = .006$; other coefficients, $p \geq .206$). Finally, telling negative lies was seen to harm the audience when they misrepresented an avoided or trustworthy agent (all $p$s $< .001$) but to help the audience when they exaggerated an untrustworthy agent ($p = .047$; other coefficients, $p$s $\geq .187$).

These findings were highly consistent with those of the main study. In particular, participants believed that truth would positively reciprocate agent behavior and negative exaggeration would negatively reciprocate agent behavior. They also believed that truth and negative exaggeration would help the audience and misrepresentation would harm the audience. Participants appeared to anticipate that out-group members would be less likely to act on their gossip than neutral and in-group members, conditioning the ability of messages to achieve their social welfare motives.

## Discussion

It is widely assumed that, given the opportunity and a selfish reason for doing so, gossipers will lie and that this means that gossip is unlikely to support the indirect reciprocity that shores up cooperation. This study suggests that there may be many circumstances in which this assumption does not hold true.

Replicating the findings of Fonseca and Peters (2018), we found that gossipers do lie (although not all gossipers and not all the time) and that self-interested motives are not the only ones in play. Thus, Feinberg, Willer, Stellar, and Keltner's (2012) finding about the importance of prosocial motives for people's sharing of gossip extends to people's decisions to distort it. Specifically, when it comes to exaggeration lies, gossipers report using them to help the audience better discriminate between targets and thereby reciprocate the target's previous behavior. In other words, it seems that exaggeration lies are the product of gossipers' attempts to actively engineer indirect reciprocity. This observation aligns with recent theorizing on the evolution of communication that suggests that in noisy environments, where audiences may miss signals, exaggeration is an expected adaptation (Wiley, 2017). Misrepresentation lies, in contrast, are the prototypical harmful lie—believed to be useful for harming the audience and, therefore, directed predominantly at competitors. The behavioral data generally support gossipers' expectations about the impact of these different lies on the welfare of the target and audience.

The above analysis points to one important caveat to the aforementioned assumption. If lies take the form of exaggeration, then there is no reason to suppose that this should erode cooperation, and indeed, it is possible that it may bolster it more than the truth. If, however, they take the form of misrepresentation, discriminating gossipers and their targets will experience paradoxical outcomes. Over time, gossipers should stop attending to gossip, and targets should stop expecting rewards for cooperative behavior. With this, the benefits of gossip should disappear. In terms of understanding the apparent robustness of cooperation in our study to the presence of lies, this presents one potential answer: The rate of misrepresentation may not have been high

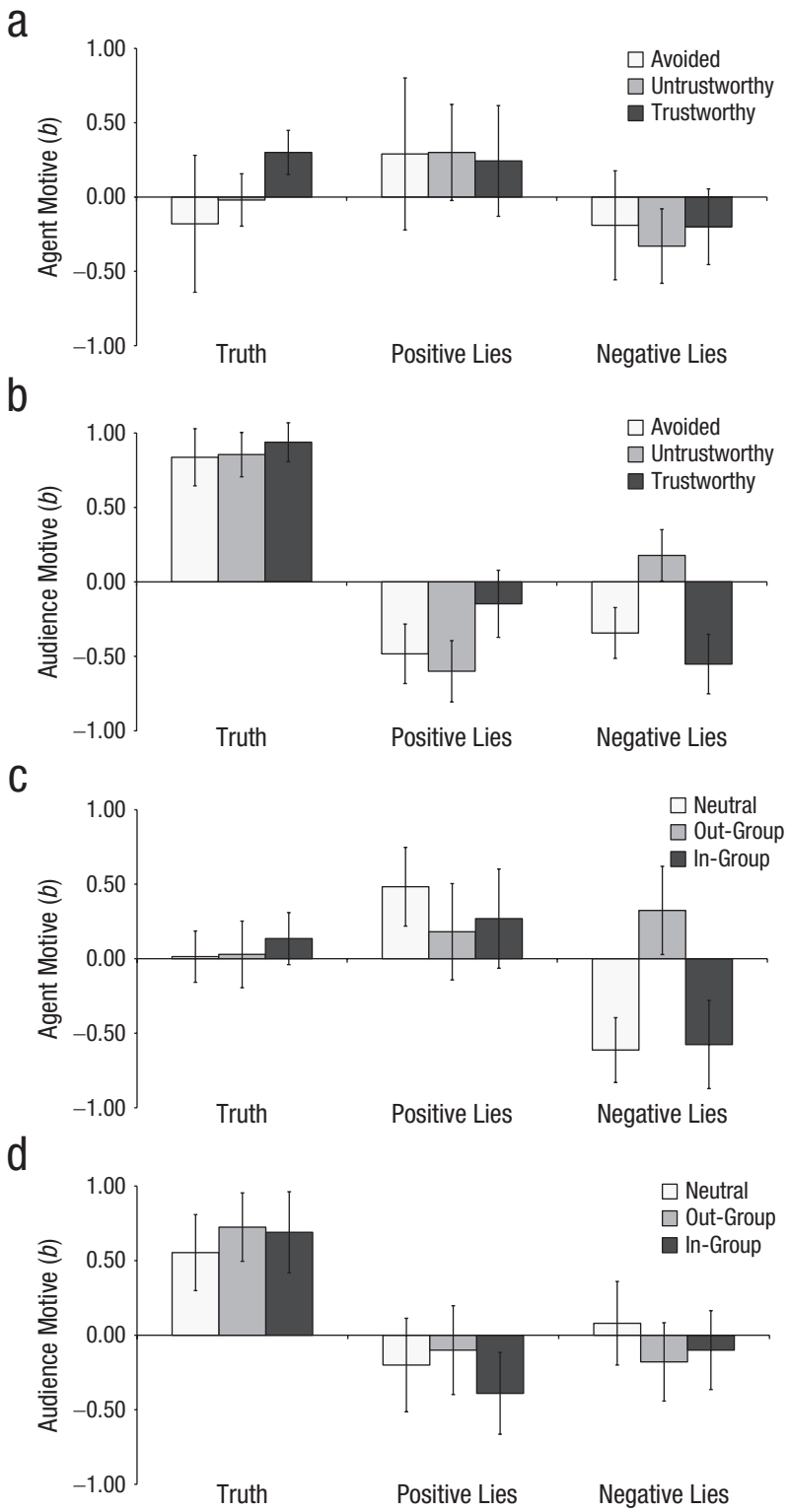


**Fig. 5.** Results of the regression on message-motive achievement as a function of the motive's target and message type. Results are shown in (a) and (c) for analyses in which an agent was the motive's target and in (b) and (d) for analyses in which the audience was the motive's target. In each panel, results for each message type are broken down by the agent's trustworthiness (a, b) or audience affiliation (c, d). Positive and negative coefficients represent message achievement of motives to help and harm, respectively. Agent-trustworthiness dummies were included in the analyses shown in (a) and (b); audience-affiliation dummies were included in the analyses shown in (c) and (d). Error bars represent 95% confidence intervals.

enough to make anomalous outcomes sufficiently frequent to produce behavior change. Indeed, targets in this study did not even appear to have insight into the levels of discrimination in their network. Together with previous work that shows that targets are quick to exploit low discrimination when it is explicitly pointed out to them (D. Fehr & Sutter, 2019; Fonseca & Peters, 2018), this suggests that the task of inferring discrimination on the basis of one's treatment is a difficult one.

These observations provide some basis for expecting that gossip may be more robust to lies in everyday life than is commonly supposed. First, it is unlikely that misrepresentation lies will dominate everyday gossip. Misrepresentation always comes with a cost: Gossipers can either mislead the audience (and thereby do them harm) or reciprocate the target's behavior; they cannot do both. In other words, the degree to which gossipers wish to harm their audience needs to be sufficiently high for them to sacrifice their powerful drive to achieve reciprocity (E. Fehr & Gächter, 2000). And second, as long as people believe that there is a sufficient chance that gossip is exchanged, is accurate, and is attended to, they may adopt a risk-averse strategy of being more cooperative than they would otherwise.

These claims are, of course, speculative and are therefore worthy of future empirical attention. It is also important to acknowledge the limitations to our claims that are attached to our study's operationalization. Specifically, our study was designed to test arguments that emerge from the literature on indirect reciprocity, which is primarily concerned with understanding cooperation in large networks that involve limited opportunities for repeated interaction. In everyday life, people belong to many networks that allow for repeated interaction with some (if not all) network members. In such networks, direct reciprocity, as well as a variety of other social motives, may come into play. Specifically, in circumstances in which gossipers are not anonymous, gossipers' desire to enhance their status and build social bonds with their audience is, among other things, likely to play an important role in their decisions to share gossip—and potentially to distort it (see Beersma & Van Kleef, 2012). How the multiplicity of motives that accompany gossip in many everyday circumstances will affect the dynamics that we describe here is an open question. It is also possible that patterns of lying that we observe for cooperative behavior (which can be expected to vary over time, whether because of deliberate defection, error, or retaliation; e.g., Charness & Rabin, 2002) may differ from behaviors or characteristics that are less variable and therefore easier for audiences to verify. Further theorizing in these domains is needed.

In sum, this research suggests that gossip can be inaccurate but that this is far from a fatal flaw. To understand how and why gossipers lie, and the social effects of those lies, we need to move beyond an assumption that these lies are necessarily malicious and harmful and consider the evident richness in their forms and functions.

## Transparency

*Action Editor:* Timothy J. Pleskac
*Editor:* D. Stephen Lindsay
*Author Contributions*

K. Peters and M. A. Fonseca contributed equally to all aspects of the study design, data collection and analysis, and manuscript writing. Both authors approved the final manuscript for submission.

*Declaration of Conflicting Interests*

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

*Open Practices*

All data and materials have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/k3jsk. The design and analysis plans were preregistered at https://osf.io/7n6qk. Note that Hypotheses 1 through 4 in the preregistration are equivalent to Hypotheses 1, 2a, 2b, and 2c, respectively, in this article. Only simplified versions of the analyses for Hypotheses 2b and 2c are reported here; full results are reported at https://osf.io/eq7n5/. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620916708. This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.


## ORCID iD

Kim Peters https://orcid.org/0000-0001-8091-8636

## Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620916708

## References

Alexander, R. D. (1987). *The biology of moral systems*. London, England: Routledge.

Barclay, P. (2012). Harnessing the power of reputation: Strengths and limits for promoting cooperative behaviors. *Evolutionary Psychology*, *10*, 868–883.

Beersma, B., & Van Kleef, G. A. (2012). Why people gossip: An empirical analysis of social motives, antecedents, and consequences. *Journal of Applied Social Psychology*, *42*, 2640–2670.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity and social history. *Games and Economic Behavior*, *10*, 122–142.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, *117*, 817–869.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, *50*, 1431–1451.

Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavior and Brain Sciences*, *16*, 681–735.

Fehr, D., & Sutter, M. (2019). Gossip and the efficiency of interactions. *Games and Economic Behavior*, *113*, 448–460.

Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, *14*, 159–181.

Feinberg, M., Willer, R., Stellar, J., & Keltner, D. (2012). The virtues of gossip: Reputational information sharing as prosocial behavior. *Journal of Personality and Social Psychology*, *102*, 1015–1030.

Fonseca, M. A., & Peters, K. (2018). Will any gossip do? Gossip need not be perfectly accurate to promote trust. *Games and Economic Behavior*, *107*, 253–281.

Giardini, F. (2012). Deterrence and transmission as mechanisms ensuring reliability of gossip. *Cognitive Processing*, *13*, 465–475.

Giardini, F., & Wittek, R. (2019). Gossip, reputation, and sustainable cooperation: Sociological foundations. In F. Giardini & R. Wittek (Eds.), *The Oxford handbook of gossip and reputation* (pp. 23–46). New York, NY: Oxford University Press.

Hess, N. H., & Hagen, E. H. (2006). Psychological adaptations for assessing gossip veracity. *Human Nature*, *17*, 337–354.

Mace, R., Thomas, M. G., Wu, J., He, Q., Ji, T., & Tao, Y. (2018). Population structured by witchcraft beliefs. *Nature Human Behaviour*, *2*, 39–44.

McAndrew, F. T., & Milenkovic, M. A. (2002). Of tabloids and family secrets: The evolutionary psychology of gossip. *Journal of Applied Social Psychology*, *32*, 1064–1082.

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature*, *393*, 573–577.

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*, 1291–1298.

Peters, K. & Kashima, Y. (2007). From social talk to social action: Shaping the social triad with emotion sharing. *Journal of Personality and Social Psychology*, *93*, 780–797.

Peters, K., & Kashima, Y. (2015). Bad habit or social good? How perceptions of gossiper morality are related to gossip content. *European Journal of Social Psychology*, *45*, 784–798.

Roberts, G. (2008). Evolution of direct and indirect reciprocity. *Proceedings of the Royal Society B: Biological Sciences*, *275*, 173–179.

Smith, E. R. (2014). Evil acts and malicious gossip: A multiagent model of the effects of gossip in socially distributed person perception. *Personality and Social Psychology Review*, *18*, 311–325.

Wedekind, C., & Milinski, M. (2000). Cooperation through image scoring in humans. *Science*, *288*, 850–852.

Wiley, R. H. (2017). How noise determines the evolution of communication. *Animal Behaviour*, *124*, 307–313.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review*, *93*, 133–138.