

جامعة نيويورك أبوظبي

 NYU | ABU DHABI



The limits of verification in preventing the spread of false information on networks

Kinga Makovi and Manuel Munoz-Herrera

Working Paper # 0038

March 2020

Division of Social Science Working Paper Series

New York University Abu Dhabi, Saadiyat Island P.O Box 129188, Abu Dhabi, UAE

<http://nyuad.nyu.edu/en/academics/academic-divisions/social-science.html>

The limits of verification in preventing the spread of false information on networks

Kinga Makovi

New York University Abu Dhabi, e-mail: km2537@nyu.edu

Manuel Muñoz-Herrera

New York University Abu Dhabi, e-mail: manumunoz@nyu.edu

Abstract

The spread of false information on social networks has garnered ample scientific and popular attention. To counteract this spread, verification of the truthfulness of information has been proposed as a key intervention. Using a behavioral experiment with over 2000 participants we analyze individuals' willingness to spread false information in a network. All individuals in the network have aligned incentives, making lying attractive, countering an explicit norm of truth-telling that we imposed. We investigate how verifying the truth, endogenously or exogenously, impacts the choices to lie or to adhere to the norm of truth-telling, compared to a setting without the possibility of verification. The three key take-aways are: (i) verification is only moderately effective in reducing the spread of lies; its effectivity is (ii) contingent on the agency of individuals to seek truth, and (iii) the exposure of liars, and not only the lies told. These suggest that verification is not a blanket solution. In order to enhance its effectivity, it should be combined with fostering a culture of truth-seeking and with information on who spreads lies, not only on the lies told.

This version: March 2020

Keywords: Verification, False information, Lying, Echo chambers, Information diffusion on networks

1 Introduction

The spread of false information on social networks has received a great deal of attention by both academic research and popular news (Vosoughi et al. 2018; Lazer et al. 2018; Ha et al. 2019). This recent interest has been sparked by the alarming potential impact that false information may have had on election outcomes Mocanu et al. (2015); Persily (2017). However, the concern applies more broadly, where false information could influence which policy to support Ding et al. (2011), whether or not one’s children should be vaccinated Schmitt et al. (2003) or if one should get a flu shot Nyhan and Reifler (2015). Against this backdrop, the study of interventions that could counteract the spread of false information on social networks is timely Iyengar and Massey (2019). A widely proposed intervention is promoting the verification of the truthfulness of information to counteract the ills that falsehood may cause (Vosoughi et al. 2018).

Verification can occur in two main ways: exogenously or endogenously. Exogenous verification is when an external and impartial source labels the veracity of information. For example, algorithms have been proposed to rank content by its credibility Ratkiewicz et al. (2011); Gupta et al. (2014). Similarly, Google has lead an effort to rank search results by a trustworthiness score Dong et al. (2015). In other words, exogenous verification is a top-down solution. Endogenous verification is when those exchanging information take measures themselves to investigate the truth. For instance, Facebook spearheaded a controversial effort of crowd-sourcing verification¹. In other words, endogenous verification is a bottom-up solution, and depends heavily on the willingness of individuals to commit resources to truth-seeking.

The underlying assumption motivating the promotion and use of verification is a widely held norm for truth-telling Abeler et al. (2019). The expectation is that when false information is identified, individuals can be expected to make it known, and not spread lies, even when it goes against their self interest. Despite this normative expectation, the effectivity of verification may be compromised when individuals do not act in a vacuum, rather, in a naturally occurring social network in which those connected to one another have similar dispositions, interests and incentives Yang et al. (2011); Colleoni et al. (2014). It has been shown that social networks have become more polarized over time Lelkes (2016); Boxell et al. (2017); Boutyline and Willer (2017); Steglich (2018), which may lead people to prioritize fitting in and supporting views that and shared by other group members, and thus beneficial to their group in reinforcing group-identity Cowan (2014); Cowan and Baldassarri (2018); Steglich (2018), instead of incorporating contradicting information Garrett et al. (2013); Becker et al. (2019), and telling the truth Abeler et al. (2019); Gneezy et al. (2018).

¹<https://www.facebook.com/zuck/videos/10106612617413491/>

This tension between aligned interests, in so-called echo chambers, and the widely held norm of truth-telling motivates our investigation of how well verification works, and how can its effectivity be enhanced. Arguably, answering this question in a field setting is fraught with challenges, for at least three reasons. First, it is nearly impossible to identify who engaged in any form of verification, impeding an evaluation on how verification impacted the choice to spread false information. Second, even if tracking this information was possible, those who verify may have different preferences for honesty than those who do not. As a consequence, it would be impossible to know if verification would be effective, if imposed, on those who do not typically verify. Third, social connections are not random, and they may depend on preferences for honesty, as well as the act of verification. In short, individuals are not randomized to social positions, nor are they randomly exposed to verification regimes, and their use.

To address these obstacles, we conduct an online experiment which provides us with a controlled environment where a verification regime can be randomly assigned and tracked. This allows us to observe which individuals know the truthfulness of the information that they spread, and the type of verification used to find out. Moreover, we have control over the social positions that individuals take, i.e., participants do not choose their interaction partners. Most importantly, our experimental design emphasizes the tension between aligned interest in ones' network, and an explicitly imposed social norm of truth-telling.

Adding to these benefits, the experiment we designed also helps us to better understand the mechanisms that may drive the effectivity of verification, such as the psychological cost that individuals experience when telling lies, or the reputational cost they perceive when identified as liars Abeler et al. (2019). We interrogate these mechanisms by introducing additional manipulations to see which of these channels may enhance the effect of verification. Our findings can help inform the design of useful interventions and policies to prevent the spread and amplification of lies on social networks in real-world settings, where people are surrounded by others who are similar to them when sharing information Weisel and Shalvi (2015); Barr and Michailidou (2017); Pennycook et al. (2019).

2 Experimental design

We designed a one-shot sequential game, which we call the *web of lies* game (see Fig. 1), where three players are assigned to different positions in a linear communication network: first, F , intermediate, I , and last, L . At the beginning of the game, player F chooses a card from a 12×12 grid, which reveals an integer, x , between 1 and 30 written on the card. The number x is observed only by F

and referred to as the *hidden number*. Player F then sends a number, x^F , also between 1 and 30, to player I , reporting on x . Player I observes x^F , but not x , and reports a number, x^I , under the same conditions to player L . Finally, player L observes x^I , but not x or x^F , and reports a final number, x^L , this time to the experimenter.

Each of the three players earn 5 cents times the number reported by the last player, i.e., if $x^L = 20$, then each of the players make \$1.00. This means that the monetary compensation is increasing in the number reported by the last player, and that the monetary incentives of individuals in the communication network are perfectly aligned. It is common knowledge that the report that players send need not to match the number they observed. This creates the possibility to over-report, i.e., the possibility to lie, and payoffs are highest when $x^L = 30$. However, all players are told that the goal of the game is for them to send reports so that the last player can report the same number that was drawn by the first player. That is, x^L should be equal to x . This rule, which is restated on the screen where players make their decisions, institute a norm of truth-telling.

We chose a distribution of hidden numbers where any integer between 1 and 30 has a positive probability to be drawn by player F , which ensures that no reports are obvious lies. As there are more cards with smaller numbers, the probability is higher for smaller numbers to be drawn which is known to all (see Supplementary Information, SI). To sum this up, the truth is costly, as far as the monetary incentives of the players are concerned, and lies may be suspected based on the size of the reports, but are never evident in short of verification. From a normative perspective, however, lies come at a cost, as players have been informed how they *should* play the game.

In this setting, to evaluate the role of verification on lying, we designed three experimental treatments that manipulate the last player's ability to observe the value of the hidden number, x , before making the final report (see Fig. 1). That is, we focalize our implementation of verification on player L , whose report is the only payoff-relevant for the group. The first treatment is a baseline with *no verification* of the hidden number (NO). That is, the game as described above, where player L makes the final report after observing x^I and has no additional information on x (see Fig. 1.4a). The second treatment introduces *exogenous verification* (EXO), such that the last player has an 80% chance of learning x after receiving x^I , but before making his report (see Fig. 1.4b). Substantively, we implement a fact checker as a centralized, external source of the veracity of information. Finally, the third is a treatment with *endogenous verification* (ENDO) where the last player, after receiving x^I , can verify x by clicking on a button (see Fig. 1.4c). This means that, unlike EXO, player L in ENDO has the choice of avoiding information by not clicking on the button. Importantly, there is no monetary cost for verifying and observing x in the verification treatments, which allows player L to identify whether the number x^I he received is a lie or not, but not who the liar is. It may

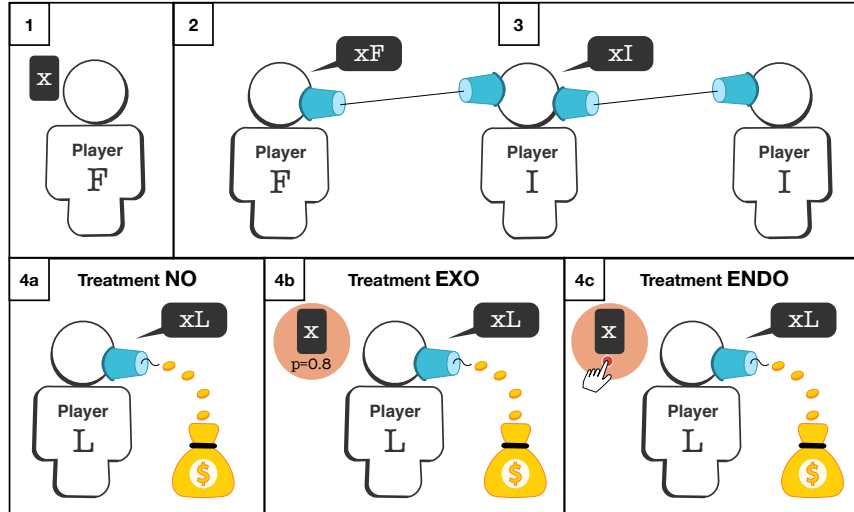


Figure 1. Experimental game and main treatment variations

Note: (1) The *web of lies* game begins with player F who draws a card with a hidden number x between 1 and 30, (2) player F then reports a number x^F to the intermediate player I , (3) player I observes x^F and reports a number x^I to the last player L , (4) player L observes x^I and reports a number x^L to the experimenter. The goal of the game is to report $x^L = x$ to the experimenter, but players are paid according to x^L , which gives them incentives to over-report. Experimental treatments manipulate the last players ability to observe the true value of x before making the final report: (4a) treatment **NO** is the baseline, where player L does not receive any information on x , (4b) treatment **EXO** is a condition with exogenous verification, where player L receives the value of x with 80% chance, before sending his report x^L . (4c) treatment **ENDO** is a condition with endogenous verification, where player L can click on a button to observe the value of x , before sending his report x^L .

either be that player F , player I , or both.

In addition, we elicited beliefs (incentivized) in both verification treatments by asking players F and I , after each made their report, whether they believed the last player had verified. That is, if L had been shown (in **EXO**) or had clicked on a button to see (in **ENDO**) the true value of x (see the full instrument in the SI). By means of this we can evaluate how the anticipation of verification impacted their behavior and incentives to lie.

The experiment was programmed in oTree (Chen et al. 2016), so that all interactions between participants took place through a web interface. We recruited participants from Amazon Mechanical Turk, which has recently gained prominence for the collection of behavioral data online Buhrmester et al. (2011); Sprouse (2011). In total, 2,177 individuals participated in nine experimental conditions, with approximately 80 groups per condition (see details below). On average, the duration of the study was 10 minutes and participants earned about \$2.0, resulting in a hourly wage of \$12.0 on average. The study was approved by the Institutional Review Board at NYU Abu Dhabi. The standard measures of anonymity and non-deception were used.

3 Results

In our analysis, we make multiple comparisons between the outcomes of the different experimental treatments using two-sided t-tests, and report p-values in the text. To foreshadow the structure of our analysis and main results, first, we quantify the effect of verification on the spread of lies at the group level and find that verification has limited effectivity in the prevention of lies. Second, we try to enhance its effectivity in additional experimental treatments using two strategies: we increase the psychological cost of lying by introducing passive players whose payoffs are reduced with false reports, and increase the reputational cost of lying by making evident who tells lies. Of these two strategies, only the latter works.

3.1 The main effect of verification on the spread of lies

To evaluate how effective verification is in reducing the spread of false reports, we focus on two key measures: the likelihood of lying, and the size of the lies told. We begin by analyzing group-level outcomes, and then turn to the behavior of individuals in each position in the communication chain.

At the group level, a lie is reported if the final report is different from the hidden number, $xL \neq x$, and the size of the lie told is the magnitude of that difference, $xL - x$ (see bars in Fig. 2A). We compare the likelihood of lies and the average size of lies of each treatment with verification to no verification. We find that only endogenous verification is effective in preventing the spread of lies, as groups in ENDO lie 22 percentage point less often ($p = 0.004$) and tell smaller lies, 7.7 versus 10.4 ($p = 0.084$), than in NO. In contrast, groups in EXO lie at the same rate ($p = 0.118$) and by an indistinguishable amount (10.4 versus 9.5 respectively, $p = 0.576$) compared to NO.

Importantly, this moderate effect on preventing lies in the treatments with verification is not due to low levels of actual verification of the truth (see diamonds in Fig. 2A). Verification was randomly assigned to a large share of groups in EXO and was chosen by an even larger share in ENDO ($75\% < 89\%$, $p = 0.020$). Even though verification in ENDO is a choice and not experimentally imposed, there is little evidence of information-avoidance. To assess what is driving the observed effectivity of endogenous verification (and the lack thereof of exogenous verification) we analyze the way reports and lies spread by respondents in the different positions in the network.

First, we consider the hidden number x , which is the value first players are asked to report on. Because x is a result of a draw in each group, and is not randomized, we compare the distribution of hidden numbers across treatments, to rule out that differences in outcomes are based on the conditions groups start in instead of our experimental manipulations (i.e., failed randomization). We find that the average x drawn, 10.6 in NO, 10.0 in EXO, and 9.7 in ENDO, are not statistically

significantly different across treatments (see first bars in Fig. 2A, as well as Table S3), indicating that groups in all three treatments start in equivalent conditions.

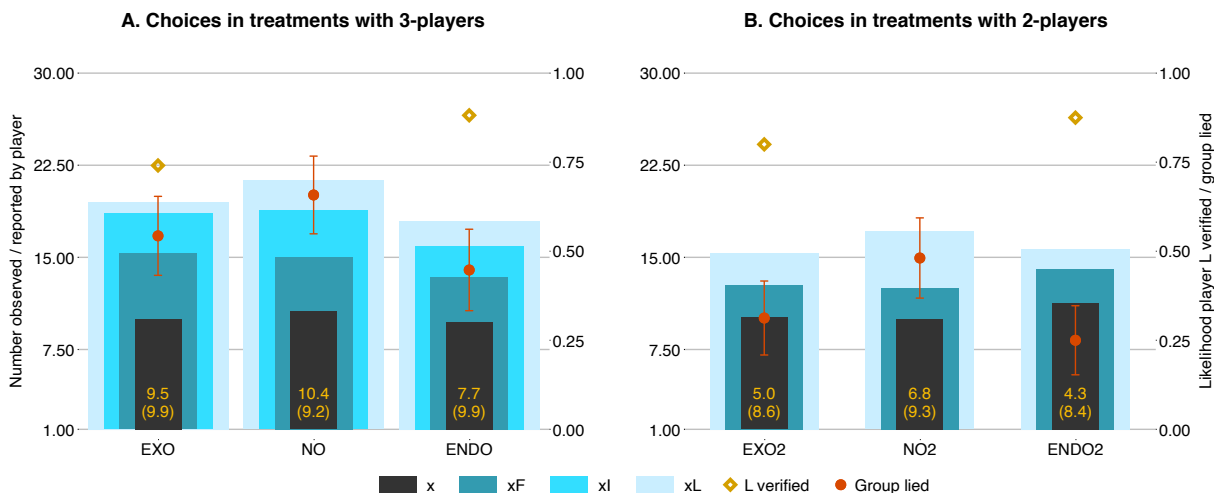


Figure 2. Main decisions by players and treatments in the 3-person games (Panel A) and in the 2-person games (Panel B).

Note: The bars indicate the numbers observed and reported by each player in the game (left vertical axis). x is the hidden number, x^F is the number reported by player F and the difference between the two bars is the magnitude of the lies told by first players; x^I is the number reported by the intermediate player I and the difference with the x^F bar is the magnitude of I 's lie; and x^L is the number reported by the last player L , where the difference between x^L and x^I is the magnitude of L 's lies, while the difference between x^L and x is the magnitude of the lie at the group level. The mean and SD of the magnitude of the group lies are at the bottom of each bar. The circle dots indicate the likelihood that a group lies in each treatment, error bars are ± 1 SE; and the diamond dots the likelihood that player L verifies (right vertical axis).

Second, we analyze the reports made by players F in the network. The average reports by players F were 15.0, 15.3, and 13.4 in NO, EXO and ENDO, respectively. Players F lie by reporting a different number than the hidden number that they drew ($x^F - x \neq 0$), and the size of their lie is the magnitude of that difference (see the gap between the first and second bars in Fig. 2A). Neither the probability of lying, or the size of the lies are affected by the anticipation of exogenous ($p = 0.788$ and $p = 0.976$) or endogenous verification ($p = 0.407$ and $p = 0.468$) when compared to the baseline condition of no verification. This suggests that knowing the last player could identify whether the number they receive as a report is false or not, does not impact the lying behavior of the first players.

Third, we consider the reports made by players in the intermediate position I . The average reports by players I were 18.7, 18.2, and 15.8 in NO, EXO and ENDO, respectively. Players I lie by reporting a different number than the one they received from player F ($x^I - x^F \neq 0$), and the size of their lie is the magnitude of that difference (see the gap between the first and second bars in Fig. 2A). Note that Players I may report a lie by repeating the report they received from player

F , if player F lied. Compared to the baseline, there are no significant differences in the magnitude of the lies told by I in ENDO ($p = 0.244$) or in EXO ($p = 0.732$). However, the probability that I lies is significantly smaller with endogenous verification ($p = 0.046$). When we analyze how the reports from the intermediate player differ from the hidden number (see the gap between the first and third bars in Fig. 2A), we find there is a significantly lower share of false reports reaching the final player in ENDO compared to NO ($p = 0.045$), moreover, smaller numbers are reported by the intermediate player $15.9 < 18.8$ ($p = 0.058$). On the contrary, no statistically significant differences are found between exogenous verification and NO in either case. That is, while in both verification conditions a downward shift is observed in lies and reports, the impact of exogenous verification cannot be distinguished from zero.

Taken together, the evidence so far suggests that players L receive different messages on average when endogenous verification is implemented compared to no verification, while no such difference exists between exogenous verification and no verification. This difference is driven by differences in the behavior of the intermediate player, possibly due to the proximity to the last player who is the one verifying. Differences in proximity may be indicative of players being weary to be found out as liars. We test this empirically by examining players beliefs, and anticipating that those who believe that they would be verified upon lie less. A total of 78% of players F and I in ENDO believed that player L would verify and observe the true value of x . When comparing the reports of those who believed that the last player would check the hidden number to those who believed the opposite, the differences are not statistically significant considering each role, F and I separately ($p > 0.100$). When we pool our data for all non-final players, we find that those who believed that the last player would verify reported significantly smaller numbers than those who believed no verification would occur ($p = 0.060$).

We finally turn to the role that last players have in stalling the spread of lies when they verified the truth. A large share of players L repeat the report they received from player I ($x_L = x_I$). However, when the message received is identified to be false, the spread of lies by repeating the report received is significantly reduced compared to the baseline, both in the probability of telling a lie and the size of the lie told (see Tab. S4 and Tab. S5). Moreover, in ENDO, where players L took agency for identifying the truth, they are more likely to correct a lie and submit smaller final reports compared to last players in EXO, where observing the true value of x was not a choice (see the gap between the third and fourth bars in Fig. 2A, also note the negative interaction terms in Tab. S6 and Tab. S7 with p-values of $p = 0.100$ and $p = 0.093$, respectively).

Our analysis reveals that endogenous verification is a more effective intervention than exogenous verification when compared to no verification. We identify two avenues via which this occurs. On

the one hand, endogenous verification reduces lies by those who anticipate that their lies will be identified. On the other, it appears to trigger the motivation to correct a lie and report the truth, by those who took agency to verify. However, our data reveals that groups still report lies in ENDO and that the reduction in the size of lies told is only moderate compared to NO. Drawing from the literature on lying behavior and the effects of transparency, the mechanisms that lie behind truth telling are multiple. The two most prominently discussed are the *psychological costs of lying* and *concerns for reputation* (Abeler et al. 2019). We explore next six new treatments that aim to elevate these costs to increase the effectivity of verification, when verification is endogenous.

3.2 Increasing the psychological cost of lying to enhance the impact of verification

First, we address how increasing the psychological cost of lying may enhance the efficacy of verification. For this, we designed two additional treatments using ENDO, where we add one or two passive players, creating victims of lies, whose payoffs decrease as a function of the lies told in the final report. We label these treatments as VCTM, when there is a single passive player, and VCTMS when there are two passive players. Victims make no decisions and their payoffs are 5 cents $\times (2x - xL)$. This means that victims earn the same as the active players when the last report is truthful ($xL = x$), but their payoffs are negatively affected by lies. This simple point is made clear to participants, stating that truth-telling results in an equal payoff for everyone. Unlike in ENDO, in VCTM and VCTMS lying by reporting a number different from the hidden number hurts others, which is expected to increase the psychological cost of lying of final players. Thus, we evaluated if the presence of negative externalities decreases the likelihood of lies, or reduces their size compared to ENDO.

We find that having a victim (VCTM) does not affect the rate at which last players verified ($p = 0.935$) compared to the victimless case of ENDO. This holds true even when the number of victims that are hurt is increased to two (VCTMS), making the size of the group that is hurt by lies equal to the size of the group that benefits from lies, disregarding player L ($p = 0.657$). Moreover, the reports made by players in all three position and the frequency and amount of lying were not different in VCTM or VCTMS compared to ENDO (see Fig. S1). In sum, we find that increasing the cost of lying by introducing negative externalities does not seem to increase the efficacy of verification. These findings are consistent with ignorance towards third-party externalities Bland and Nikiforakis (2015), and triggering a rational mode of thinking, which encourages cheating Amir et al. (2016).

3.3 Increasing the reputational cost of lying to enhance the impact of verification

Next, we test whether the effectivity of verification can be enhanced by increasing the reputational costs of lying of non-final players. We expect that players would want to avoid being seen as liars by their group members. We found indicative evidence of this conjecture when we observed that the expectation of verification had a stronger effect on players who believed that they would be verified upon, as well as that players positioned closer to the one verifying lied less. To enhance the salience of the reputational mechanism, we designed three additional experimental treatments where not only the lie, but also the liar can be identified through verification. We achieve this by removing the intermediate player I from the communication chain, which means that we now analyze a 2-player *web of lies* game.

In this setting, we test the role of verification, as we did before, comparing a treatment with no verification (NO2), to two forms of verification: exogenous verification (EXO2) and endogenous verification (ENDO2). In this 2-player game, the first player F observes x and reports a number to the last player L . Therefore, if F tells a lie, she risks being identified as a liar and cannot hide behind another group member who may have also lied. This is true even in NO2 where L has no certainty of F being a liar, but is aware of the distribution from which x was drawn (note that in the 3-player game the likelihood of a lie was also identifiable but not the source of the lie). In other words, diffusing responsibility for a lie is not possible Conrads et al. (2013). As a corollary, reputational concerns are high and virtually equal across experimental conditions.

Eliminating the intermediate player, and thus the possibility of hiding behind others, results in first players telling fewer and smaller lies than those in the same position in the main treatments with 3-players, in NO2 ($2.6 < 4.8, p = 0.058$), EXO2 ($2.6 < 5.2, p = 0.034$) and ENDO2 ($2.8 < 3.8, p = 0.393$ —note that although the effect is present it is significant in ENDO2, probably due to floor effects). Moreover, individuals in the position of player F do not differ in their behavior across treatments in the 2-person game either (see Fig. 2B). This provides an ideal test on the effect of verification, given that players L observe statistically indistinguishable messages (and lies) when they make their reports. Our findings show that at the group level, removing the intermediate player leads to less frequent and smaller lies than in the 3-person treatments (NO2 $p = 0.016$, EXO2 $p = 0.002$ and ENDO2 $p = 0.020$).

Consistent with our previous findings, the effect in the conditions with verification (EXO2 and ENDO2) is not driven by differences in the rates of verification, which were not different across communication chain lengths (see diamonds in Fig. 2B). On top of that, evaluating the effect of verification we find that both EXO2 and ENDO2 are effective in reducing the share of lying groups

compared to NO2: $p = 0.033$ and $p = 0.002$, respectively (see circles in Fig. 2B). However, the same effect is only marginally significant in reducing the size of lies for ENDO2 ($p = 0.078$) and not significant for EXO2 ($p = 0.209$). This supports our finding suggesting that the choice to verify, and not simply observing the true value of x , is strongly driving truth-telling.

Our findings from the treatments with increased exposure of the liar, provide new insights on existing experimental evidence arguing that higher scrutiny has little impact on dishonesty. Although it has been suggested that transparency policies are not always a remedy against dishonesty, we find that being exposed as a liar to a group member, as in our case, has a significantly different effect than being exposed to other third parties, which proves ineffective van de Ven and Villeval (2015) or to the experimenter, which can even be counterproductive Gneezy et al. (2018).

Finally, we conducted an additional treatment, to interrogate if increasing both reputational and psychological costs have a compound effect. Namely, on the 2-person game we introduce a passive victim to the endogenous verification case (VCTM2) and find that the effectivity of verification did not increase ($p = 0.992$). Thus, on top of increasing reputational costs by eliminating the intermediate player, there is no additional effect on verification when introducing negative externalities from lying.

We conclude, that reputational mechanisms are key to promote the effectivity of verification, while enhancing the psychological cost of lying does not appear to be effective.

3.4 Limitations

While this work has some key advantages, it is not without limitations. Its very strength in our experimental approach results in some weaknesses. Importantly, in real world settings centralized fact-checkers do not only tag lies, but with those tags come the reputation of the institution and the true or presumed ideologies of that institution. In our experiment we can not speak to how individuals' behaviors would change as a function of this additional signal. Modification of the experiment to incorporate these nuances are possible, but are left to future work. Moreover, in this experiment all participants' information about the possible states of the world was the same, i.e., nobody was more or less informed, or had varying levels of expertise to probe what may be true, which clearly is not the case in reality. This latter point may mask important heterogeneity about the willingness to spread lies by those who are more knowledgeable or informed in the substantive domain concerning the piece of information they receive, their group-identity notwithstanding. Similarly to the above mentioned point, this is another important avenue for future research.

4 Conclusions

Our work highlights some major limitation of verification interventions on preventing the spread of false information. The spread of lies on networks that are organized around shared interests, such as the case in echo-chambers, is difficult to prevent. Even when an explicit norm of truth-telling is instituted or when lies can affect third-parties outside the group, verification is only moderately effective. This is so because simply revealing that lies spread is insufficient for stopping individuals benefiting themselves and their group members, which may be a legitimizing force behind spreading lies in such settings. However, when people take agency to verify the truth of information, it is this seeking behavior that enhances adhering to the norm of telling the truth (see Scheufele and Krause 2017).

In large networks, simply introducing verification is unlikely to work Jun et al. (2017). That is, top-down solutions are likely to fail as these centralized measures may be deemed “inconsequential” from the perspective of the group and its members. In large social networks it is easy to hide behind others, and detrimental outcomes can result without clear culprits. Taken together, an uphill battle needs to be fought to be effective in countering the spread of false information: changing the culture that values truth and tracking the source or sources where lies originate so that liars can be tagged.

Although this can be challenging, especially in a “post-truth” society, our work provides insights on how to tackle this problem. As evidenced by our results, sharing common interests make groups strong and facilitates the spread of lies. But, it is this same drive of caring about one’s connections what can be used to strengthen verification. Specifically, we observe that the potential of being exposed as a liar to those one is tied to, significantly increases adherence to the truth-telling norm. Therefore, verification strategies works best when not only lies but liars are tagged, as individuals aim to maintain a good reputation.

References

- Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4):1115–1153.
- Amir, A., Kogut, T., and Bereby-Meyer, Y. (2016). Careful cheating: People cheat groups rather than individuals. *Frontiers in Psychology*, 7:371.
- Barr, A. and Michailidou, G. (2017). Complicity without connection or communication. *Journal of Economic Behavior & Organization*, 142:1–10.
- Becker, J., Porter, E., and Centola, D. (2019). The wisdom of partisan crowds. *Proceedings of the National Academy of Sciences*, 116(22):10717–10722.

- Bland, J. and Nikiforakis, N. (2015). Coordination with third-party externalities. *European Economic Review*, 80:1–15.
- Boutyline, A. and Willer, R. (2017). The social structure of political echo chambers: Variation in ideological homophily in online networks. *Journal of Political Psychology*, 38(3):551–569.
- Boxell, L., Gentzkow, M., and Shapiro, J. M. (2017). Greater internet use is not associated with faster growth in political polarization among us demographic groups. *Proceedings of the National Academy of Sciences*, 114(40):10612–10617.
- Buhrmester, M., Kwang, T., and Gosling, S. (2011). Amazon’s mechanical turk – a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5.
- Chen, D. L., Schonger, M., and Wickens, C. (2016). otree – an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.
- Colleoni, E., Rozza, A., and Arvidsson, A. (2014). Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data. *Journal of Communication*, 64(2):317–332.
- Conrads, J., Irlenbusch, B., Rilke, R. M., and Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology*, 34:1–7.
- Cowan, S. K. (2014). Secrets and misperceptions: The creation of self-fulfilling illusions. *Sociological Science*, 1:466–492.
- Cowan, S. K. and Baldassarri, D. (2018). “It could turn ugly”: Selective disclosure of attitudes in political discussion networks. *Social Networks*, 52:1–17.
- Ding, D., Maibach, E. W., Zhao, X., Roser-Renouf, C., and Leiserowitz, A. (2011). Support for climate policy and societal action are linked to perceptions about scientific agreement. *Nature Climate Change*, 1:462–466.
- Dong, X. L., Gabrilovich, E., Murphy, K., Dang, V., Horn, W., Lugaresi, C., Sun, S., and Zhang, W. (2015). Knowledge-based trust: Estimating the trustworthiness of web sources. In *Proceedings of the VLDB Endowment*, volume 8, pages 938–949.
- Garrett, R. K., Carnahan, D., and Lynch, E. K. (2013). A turn toward avoidance? selective exposure to online political information, 2004–2008. *Political Behavior*, 35(1):113–134.
- Gneezy, U., Kajackaite, A., and Sobel, J. (2018). Lying aversion and the size of the lie. *American Economic Review*, 108(2):419–453.
- Gupta, A., Kumaraguru, P., Castillo, C., and Meier, P. (2014). Tweetcred: Real-time credibility assessment of content on twitter. In Aiello, L. M. and McFarland, D., editors, *Social Informatics*, chapter 10, pages 228–243. Springer, Berlin.
- Ha, L., Perez, L. A., and Ray, R. (2019). Mapping recent development in scholarship on fake news and misinformation, 2008 to 2017: Disciplinary contribution, topics, and impact. *American Behavioral Scientist*, 0(0):0002764219869402.
- Iyengar, S. and Massey, D. S. (2019). Scientific communication in a post-truth society. *Proceedings of the National Academy of Sciences*, 116(16):7656–7661.

- Jun, Y., Meng, R., and Johar, G. V. (2017). Perceived social presence reduces fact-checking. *Proceedings of the National Academy of Sciences*, 114(23):5976–5981.
- Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., and Zittrain, J. L. (2018). The science of fake news: Addressing fake news requires a multidisciplinary effort. *Science*, 359:1094–1096.
- Lelkes, Y. (2016). Mass polarization: Manifestations and measurements. *Public Opinion Quarterly*, 80(S1):392–410.
- Mocanu, D., Rossi, L., Zhang, Q., Karsai, M., and Quattrociocchi, W. (2015). Collective attention in the age of (mis)information. *Computers in Human Behavior*, 51, Part B(1):1198–1204.
- Nyhan, B. and Reifler, J. (2015). Does correcting myths about the flu vaccine work? an experimental evaluation of the effects of corrective information. *Vaccine*, 33(3):459–464.
- Pennycook, G., Bear, A., Collins, E., and Rand, D. G. (2019). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, page forthcoming.
- Persily, N. (2017). The 2016 u.s. election: Can democracy survive the internet? *Journal of Democracy*, 28(2):63–76.
- Ratkiewicz, J., Conover, M., Goncalves, B., Flammini, A., and Menczer, F. (2011). Detecting and tracking political abuse in social media. In *In Proceedings of the 5th AAAI International Conference on Weblogs and Social Media (ICWSM’11)*.
- Scheufele, D. A. and Krause, N. M. (2017). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences*, 116(16):7662–7669.
- Schmitt, H.-J., Booy, R., Weil-Olivier, C., Damme, P. V., Cohene, R., and Peltola, H. (2003). Child vaccination policies in europe: a report from the summits of independent european vaccination experts. *The Lancet Infectious Diseases*, 3(2):103–108.
- Sprouse, J. (2011). A validation of amazon mechanical turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods*, 43(1):155–167.
- Steglich, C. (2018). Why echo chambers form and network interventions fail: Selection outpaces influence in dynamic networks.
- van de Ven, J. and Villeval, M. C. (2015). Dishonesty under scrutiny. *Journal of the Economic Science Association*, 1:86–99.
- Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359:1146–1151.
- Weisel, O. and Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences*, 112(34):10651–10656.
- Yang, S.-H., Long, B., Smola, A., Sadagopan, N., Zheng, Z., and Zha, H. (2011). Like like alike: joint friendship and interest propagation in social networks. In *Proceedings of the 20th international conference on World wide web*, pages 537–554.

Appendix A Participants' descriptive statistics

Table A1. Descriptive statistics for treatments in the 3-player games.

Variable	NO	EXO	ENDO	VCTM	VCTMS
N of groups	80	83	80	83	77
N of participants	240	249	240	332	385
Gender (%)					
Female	48.3	51.0	47.5	50.3	50.4
Male	51.3	47.8	52.1	49.4	48.8
Other	0.4	1.2	0.4	0.3	0.8
Race (%)					
Non-Hispanic White	78.8	84.7	80.4	79.2	80.8
Other than non-Hispanic White	21.2	15.3	19.6	20.8	19.2
Education (%)					
High school or less	6.3	10.0	10.8	9.9	7.5
Some college	30.8	32.1	33.3	26.8	30.1
College diploma	47.1	44.6	42.1	48.5	45.7
More than college degree	15.8	13.3	13.8	14.8	16.6
Income (%)					
Less than \$10,000	13.8	15.3	13.8	16.0	13.8
\$10,000 - \$19,999	10.0	6.4	12.9	9.0	9.1
\$20,000 - \$29,999	12.1	14.5	11.2	11.1	13.5
\$30,000 - \$39,999	14.6	14.1	14.6	11.1	11.7
\$40,000 - \$49,999	12.5	9.2	11.7	10.5	13.5
\$50,000 - \$69,999	17.9	18.1	16.2	20.8	18.7
More than \$70,000	19.2	22.5	19.6	21.4	19.7
Age (mean years)	34.8	35.3	34.1	33.8	34.5
Compensation (mean \$ of active participants)	\$2.04	\$1.98	\$1.88	\$1.96	\$1.91

Table A2. Descriptive statistics for treatments in the 2-player games.

Variable	NO2	EXO2	ENDO2	VCTM2
N of groups	81	82	81	81
N of participants	162	164	162	243
Gender (%)				
Female	51.9	54.9	48.8	50.2
Male	48.1	45.1	51.2	49.4
Other	0.0	0.0	0.0	0.4
Race (%)				
Non-Hispanic White	82.1	75.0	84.0	78.2
Other than non-Hispanic White	17.9	25.0	16.0	21.8
Education (%)				
High school or less	7.4	7.3	7.4	11.1
Some college	28.4	32.3	30.9	26.3
College diploma	49.4	47.0	46.3	45.0
More than college degree	14.8	13.4	15.4	17.7
Income (%)				
Less than \$10,000	9.9	15.9	11.7	14.4
\$10,000 - \$19,999	13.0	9.8	4.9	9.5
\$20,000 - \$29,999	11.1	11.6	16.1	14.4
\$30,000 - \$39,999	16.1	14.0	14.2	14.4
\$40,000 - \$49,999	11.7	11.0	11.1	8.6
\$50,000 - \$69,999	17.3	18.3	20.4	18.1
More than \$70,000	21.0	19.5	21.6	20.6
Age (mean years)	35.0	34.7	35.0	35.0
Compensation (mean \$ of active participants)	\$1.85	\$1.76	\$1.78	\$1.76

Appendix B Comparison of the hidden numbers drawn

Table B1. Comparing differences in means in the hidden number across conditions.

	EXO	ENDO	NO2	EXO2	ENDO2	VCTM	VCTMS
NO	0.569	0.430	0.764	0.659	0.343	0.522	0.692
EXO		0.822	0.769	0.903	0.138	0.214	0.337
ENDO			0.598	0.731	0.092	0.140	0.240
NO2				0.871	0.209	0.326	0.480
EXO2					0.175	0.272	0.407
ENDO2						0.703	0.570
VCTM							0.825

Appendix C Regression

Table C1. Modeling the probability of lying at the group level comparing no to the verification conditions exo and endo using logistic regression.

	Estimate	Std. Error	Pr(> z)
Intercept	-0.773	0.349	0.027*
Verification	-0.062	0.429	0.886
Received a lie	4.487	1.071	<0.001***
Verification × Received a lie	-3.037	1.134	0.007**

Table C2. Modeling the final report at the group level comparing no to the verification conditions exo and endo using OLS regression.

	Estimate	Std. Error	Pr(> z)
Intercept	15.342	1.391	<0.001***
Verification	-0.118	1.704	0.945
Received a lie	10.372	1.920	<0.001***
Verification × Received a lie	-4.473	2.438	0.068.

Table C3. Modeling the probability of lying at the group level comparing exo to endo using logistic regression.

	Estimate	Std. Error	Pr(> z)
Intercept	-1.056	0.411	0.010*
Endogenous	0.363	0.518	0.484
Received a lie	2.112	0.581	<0.001***
Endogenous × Received a lie	-1.265	0.769	0.100

Table C4. Modeling the final report at the group level comparing exo to endo using OLS regression.

	Estimate	Std. Error	Pr(> z)
Intercept	14.419	1.599	<0.001***
Endogenous	1.358	2.078	0.515
Received a lie	8.516	2.262	<0.001***
Endogenous × Received a lie	-5.332	3.151	0.093

Appendix D Additional Figures

Figure D1 reports the main decisions for treatments with passive victims, compared to treatments with endogenous verification, both for the 3-person and the 2-person games.

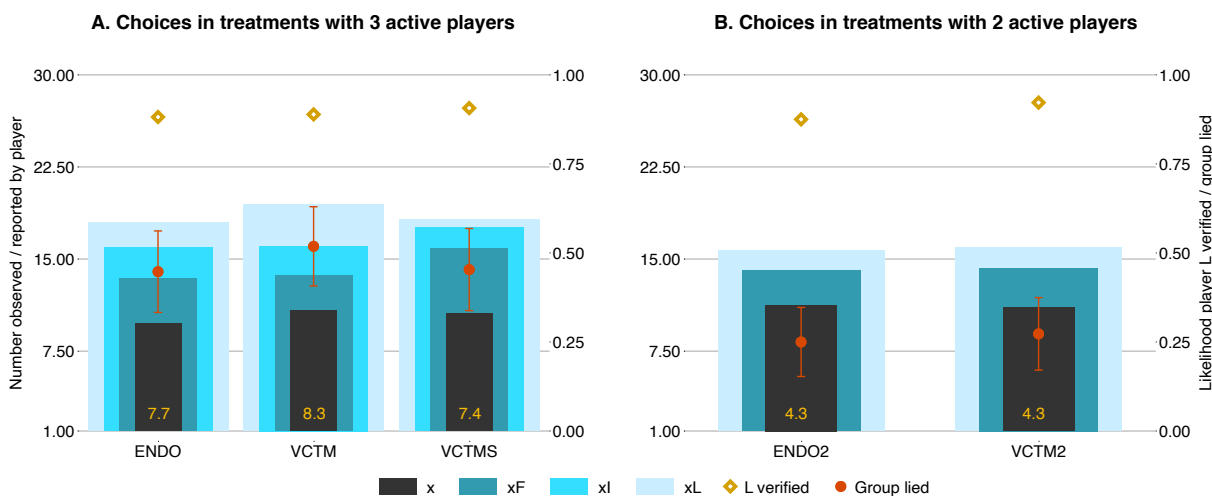


Figure D1. Main decisions by players and treatments in the 3-person games (Panel A) and in the 2-person games (Panel B) with victims.

Note: The bars indicate the numbers observed and reported by each player in the game (left vertical axis). x is the hidden number, x^F is the number reported by player F and the difference between the two bars is the magnitude of the lies told by first players; x^I is the number reported by the intermediate player I and the difference with the x^F bar is the magnitude of I 's lie; and x^L is the number reported by the last player L , where the difference between x^L and x^I is the magnitude of L 's lies, while the difference between x^L and x is the magnitude of the lie at the group level. The mean and SD of the magnitude of the group lies are at the bottom of each bar. The circle dots indicate the likelihood that a group lies in each treatment, error bars are ± 1 SE; and the diamond dots the likelihood that player L verifies (right vertical axis).

Appendix E Instructions

PAGE 1

Welcome to this study. You will play a simple game with other MTurk workers, the details of which will be explained to you in the following screens. You will read the instructions for the tasks and be asked a few comprehension questions to see if you understand the instructions. You can only participate if you answer all of these questions correctly, so your full attention is needed for the duration of the study.

Your participation will last between 5 and 10 minutes.

For participating in the study, you will earn a show-up payment of \$1. You can earn up to \$2.7 in these tasks in addition to your show-up payment. How to earn this money is explained to you in the instructions.

Please, click on the button to begin your participation in this study.

PAGE 2

Sociodemographic Questionnaire

Please answer the questions below.

1. What is your gender?
 - Male
 - Female
 - Other

2. What is your Ethnicity?
 - Hispanic / Latino / Latina
 - Not Hispanic / Latino / Latina

3. What is your Race?
 - White
 - Black / African American
 - Asian
 - American Indian / Alaskan Native
 - Middle Eastern / North African

- Native Hawaiian / Pacific Islander
 - Other
4. What is your age?
5. What is your highest completed level of education?
- Less than high school
 - High school diploma or equivalent (e.g., GED)
 - Some college
 - College diploma
 - Masters degree
 - Professional post-secondary degree or doctoral degree (e.g., JD, MD, PhD etc.)
6. Which state do you live in?
7. What was your yearly personal income in 2017 (include salary, interests, returns on investments)?
- Less than \$10,000
 - \$10,000 – \$19,999
 - \$20,000 – \$29,999
 - \$30,000 – \$39,999
 - \$40,000 – \$49,999
 - \$50,000 – \$59,999
 - \$60,000 – \$69,999
 - \$70,000 – \$79,999
 - \$80,000 – \$99,999
 - \$100,000 – \$119,999
 - \$120,000 – \$149,999
 - \$150,000 – \$199,999
 - \$200,000–
 - I do not wish to report my income.

PAGE 3

Below we present the instructions for treatment NO with 3 active players. In the following subsections we illustrate differences in instructions for treatments EXO, ENDO and VCTM. The instructions for the 2-person games are as the ones we present, without the information for the intermediate player.

Please read the instructions carefully. After you finish reading them you will be asked some comprehension questions to verify that you understand the instructions. You can only participate in the study once you have answered all of the questions correctly. If you did not get all questions right, you will have the opportunity to view the instructions again, and revise your answers one more time. **If you fail twice you will not be able to participate in this study, and you will not get paid.** As a consequence, your full attention is required for the duration of the study.

Instructions. You will earn money based on the decisions you and others make in this game explained below. **This payment is over and above the payment of \$1 dollar that you will be paid for this HIT.**

The first task you will complete is explained here, and you will receive instructions for any additional tasks later. In the first task, you will play a game in a group with two other participants who are also workers on Amazon Mechanical Turk.

You will be labeled as P1, P2, or P3, which will indicate the order in which you play. Depending on your label, you will be the one drawing the **hidden number**, you will send a message to the next player, or submit a **final message**.

The objective of the game is to send messages so that P3 would report in the **final message** the **hidden number** drawn by P1.

The task of the different participants is illustrated in detail in the image, and is described below:

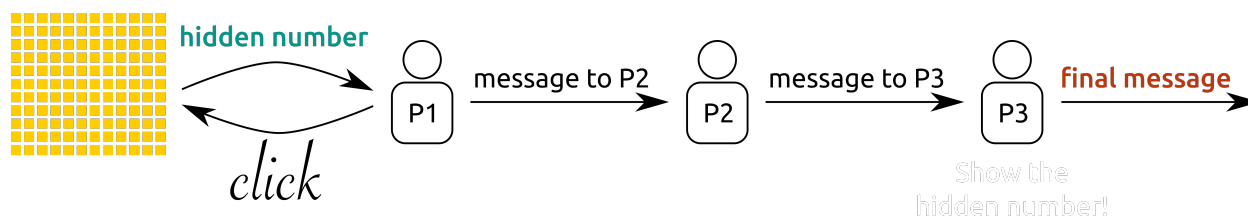


Figure D2. Decisions in the study

1. The game starts with P1 who clicks on one of the 144 cards arranged in a 12 by 12 grid appearing on the screen which can be seen on the right of the image below. Each card has a number on it between 1 and 30, and clicking on a card reveals the number, which will be the **hidden number** in the game you play.

The frequency of each number on the cards is illustrated on the left of the image at the bottom of these instructions. Note that each of the possible numbers appear on at least one card. However, there are some numbers that are more frequent than others. For example, there are four cards with the number 5 on them, there are sixteen cards with the number 7 on them, or thirty cards with the number 10 on them - which is the most frequent number -, while there is only one card with, say, the number 20 on it.

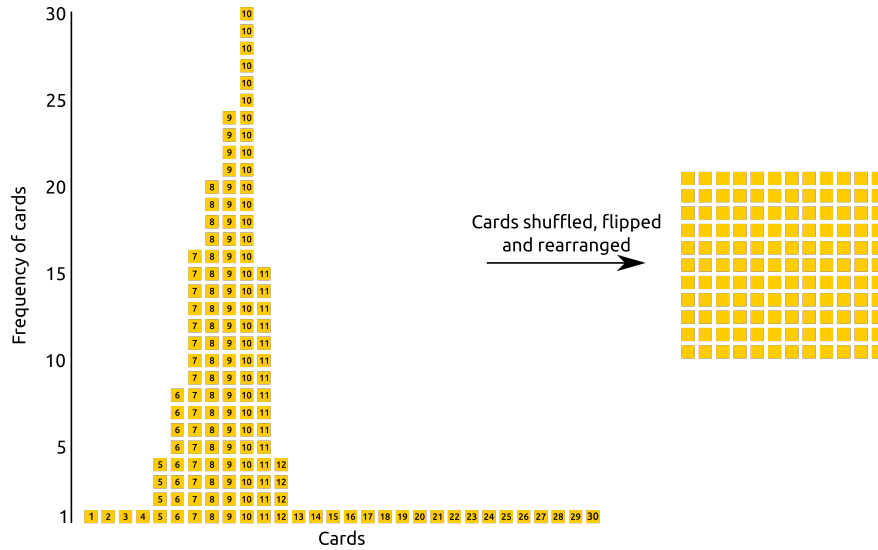


Figure D3. Distribution of the hidden number

2. The **hidden number** is revealed to P1 when he/she clicks on a card. Then, P1 will send a message reporting on the **hidden number** to P2.
3. P2 will see the number P1 reported in his/her message, but will not see the **hidden number** drawn by P1. Then, P2 will send a message to P3 reporting on the **hidden number**.
4. P3 will see the number P2 reported in his/her message. P3 will not see the **hidden number** that was drawn by P1. Then, P3 sends a **final message** reporting on the **hidden number**.

Earnings

Note that the earnings in this task are additional to your payment of \$1 dollar that you will be paid for this HIT regardless of the choices you and others make in this game.

All participants get paid according to the number P3 reports in the **final message**.

P1, P2 and P3 each earns 5 cents times the number P3 reports in the **final message**. For example, if the number reported by P3 is 16, then P1, P2 and P3 will earn $5 \cdot 16 = 80$ cents.

These instructions will be displayed at the bottom of your screen if you click on a “Show Instructions” button.

If you have read the instructions **carefully**, press the button below. You can only participate in the study once you have answered correctly **all** of the questions testing your comprehension of the instructions, on the following screen. If you do not get all questions right, you will have an opportunity to review the instructions again, and revise your answers one more time. **If you fail twice, you will not be able to participate in this study, and you will not get paid.** As a consequence, your full attention is required for the duration of the study.

PAGE 3 (other treatments)

Treatment EXO

The instructions for treatment EXO differ from NO in point #4, as follows:

4. P3 will see the number P2 reported in his/her message. P3 will also see the **hidden number** that was drawn by P1 with 80% chance. This means that with 20% chance P3 will not see the **hidden number**. Then, P3 sends a **final message** reporting on the **hidden number**.

Treatment ENDO

The instructions for treatment ENDO differ from NO in point #4, as follows:

4. P3 will see the number P2 reported in his/her message. P3 can also check the **hidden number** that was drawn by P1 by clicking on a button displayed on his/her screen. P3 will not see the **hidden number** if he/she does not click on the “Show the hidden number” button. Then, P3 sends a **final message** reporting on the **hidden number**.

Treatment VCTM

The instructions for treatment VCTM differ from NO in point #4, has an addition point #5 and a different description of earnings, as follows:

4. P3 will see the number P2 reported in his/her message. P3 can also check the **hidden number** that was drawn by P1 by clicking on a button displayed on his/her screen. P3 will not see the **hidden number** if he/she does not click on the “Show the hidden number” button. Then, P3 sends a **final message** reporting on the **hidden number**.

5. P4 does not receive or send any messages in this task, he/she is a passive player.

Earnings

All four participants get paid according to the number P3 reports in the **final message**, but the payment will be calculated differently for P4 than for the other three players (P1, P2, and P3).

P1, P2 and P3 each earns 5 cents times the number P3 reports in the **final message**. For example, if the number reported by P3 is 16, then P1, P2 and P3 will earn $5*16 = 80$ cents.

P4, on the other hand, will be paid 5 cents times the difference between twice the **hidden number** and the **final message** – which is $(2*\text{hidden number} - \text{final message})*5$ cents. For example, if the **hidden number** is 10 and P3 reported 16 in the **final message**, then P4 will earn $(2*10-16)*5 = (20-16)*5 = 4*4 = 20$ cents.

If the resulting earnings for P4 would be negative, he/she earns 0 cents in this task.

Note that if the **hidden number** and the **final message** are the same, all players earn the same amount.

PAGE 4

Comprehension Questionnaire

Please answer these questions carefully. You can only participate in the study once you have answered all of them correctly. If you answer one or more incorrectly, you will get a chance to review your answers again for a second time. If you have any doubts, you can click on the "Show Instructions" button below and the instructions will be displayed.

1. **P1 will draw a hidden number between 1 and 30 from the cards displayed in the instructions. Which of these numbers is the most likely to be drawn as the hidden number?**
 - 10
 - 4
 - 25
 - 18
 - 20
2. **P1 will send a message reporting on the hidden number to P2. What information will P2 receive?**
 - P2 will see the number reported by P1, and with 80% chance will also see the hidden number

- P2 will see the number reported by P1 and will see the hidden number if he/she clicks on a button so that it shows on the screen
 - P2 will see the number reported by P1 but not the hidden number
 - P2 will not see the number reported by P1, only the hidden number
 - P2 will not see any messages, he/she is a passive player
3. **P2 will send a message reporting on the hidden number to P3. What information will P3 receive?**
- P3 will see the number reported by P2, and with 80% chance will also see the hidden number
 - P3 will see the number reported by P2 and will see the hidden number if he/she clicks on a button so that it shows on the screen
 - P3 will see the number reported by P2 but not the hidden number
 - P3 will not see the number reported by P2, only the hidden number
 - P3 will not see any messages, he/she is a passive player
4. **Suppose the hidden number is 12 and P3 reports 12 in the final message. How many cents do participants P1, P2 or P3 earn aside from their show-up payment?**
- $24 \times 5 = 120$ cents
 - $6 \times 5 = 30$ cents
 - $12 \times 5 = 60$ cents
 - $18 \times 5 = 90$ cents
 - It cannot be determined from this information alone
5. **Suppose the hidden number is 12 and P3 reports 18 in the final message. How many cents do participants P1, P2 or P3 earn aside from their show-up payment?**
- $24 \times 5 = 120$ cents
 - $6 \times 5 = 30$ cents
 - $12 \times 5 = 60$ cents
 - $18 \times 5 = 90$ cents
 - It cannot be determined from this information alone

Belief Questionnaire

At the end of the study, before receiving information about the final report or their earnings, participants responded to a series of belief elicitation questions as follows:

Instructions

In this task you can earn **extra money** by answering a few questions. We will ask you to report how you think **other participants** played in this game you just participated in.

You will be asked a few questions in this task and **one of them** will be randomly selected for payment.

You will earn **20 cents** depending on the accuracy of your answer for the randomly selected question.

Player 1

- This question is about the choice made by another participant **in your group**. You can earn **20 cents** if your report is 3 units above or below the choice made by the other participant if this question is selected for payment. For example, if the participant reported 5, and your answer is between 2 and 8, you will earn 20 cents. Otherwise, you will not earn money in this task.

The message you sent to P2 reporting on the **hidden number** was [*report from P1 was displayed here*]. What number do you think P2 sent to P3 after seeing your report?

- EXO: This question is about the information observed by participant **P3 in your group**. If your answer is accurate, you will earn **20 cents** in case this question is selected for payment.

In addition to seeing P2's report, there is an 80% chance that P3 would see the **hidden number** drawn. Do you think P3 observed the **hidden number**?

- ENDO: This question is about the choice made by participant **P3 in your group**. If your answer is accurate, you will earn **20 cents** in case this question is selected for payment

In addition to seeing P2's report, P3 could click on a button to see the **hidden number** drawn. Do you think P3 clicked on the button and observed the **hidden number**?

- This question is about the information observed by another participant **in your group**. You can earn **20 cents** if your report is 3 units above or below the information observed by the

other participant. For example, if the participant observed 5 and your answer is between 2 and 8, you will earn 20 cents. Otherwise, you will not earn money in this task.

You said that the message P2 sent to P3 reporting on the **hidden number** was [*belief reported by P1 to the first question was displayed here*]. What number do you think P3 sent in the **final message**?

Player 2

- This question is about the information observed by another participant **in your group**. You can earn **20 cents** if your report is 3 units above or below the information observed by the other participant. For example, if the participant observed 5 and your answer is between 2 and 8, you will earn 20 cents. Otherwise, you will not earn money in this task.

The message P1 sent to you reporting on the **hidden number** was [*report from P1 was displayed here*]. What number do you think P1 drew as the **hidden number** before sending you his report?

- EXO: This question is about the choice made by participant **P3 in your group**. If your answer is accurate, you will earn **20 cents** in case this question is selected for payment

In addition to seeing your report, there is an 80% chance that P3 would see the **hidden number** drawn. Do you think P3 observed the **hidden number**?

- ENDO: This question is about the information observed by participant **P3 in your group**. If your answer is accurate, you will earn **20 cents** in case this question is selected for payment.

In addition to seeing your report, P3 could click on a button to see the **hidden number** drawn. Do you think P3 clicked on the button and observed the **hidden number**?

- This question is about the choice made by another participant **in your group**. You can earn **20 cents** if your report is 3 units above or below the choice made by the other participant if this question is selected for payment. For example, if the participant reported 5, and your answer is between 2 and 8, you will earn 20 cents. Otherwise, you will not earn money in this task.

The message you sent to P3 reporting on the **hidden number** was [*belief reported by P2 to the first question was displayed here*]. What number do you think P3 sent in the **final message**?

Player 3

- This question is about the information observed by another participant **in your group**. You can earn **20 cents** if your report is 3 units above or below the information observed by the other participant. For example, if the participant observed 5 and your answer is between 2 and 8, you will earn 20 cents. Otherwise, you will not earn money in this task.

The message you received from P2 reporting on the **hidden number** was [*report from P2 was displayed here*]. What number do you think P2 received from P1 before sending you this message?

- **Only if P3 did not verify the true value of the hidden number**. This question is about the information observed by another participant **in your group**. You can earn **20 cents** if your report is 3 units above or below the information observed by the other participant. For example, if the participant observed 5 and your answer is between 2 and 8, you will earn 20 cents. Otherwise, you will not earn money in this task.

You said the message P1 sent to P2 reporting on the **hidden number** was [*belief reported by P3 to the first question was displayed here*]. What number do you think P1 drew as the **hidden number** before sending his report to P2?

Appendix F Decision Screens

Treatment NO

In Figure F1 we display the decision screens for all three players (P1, P2 and P3) in treatment NO.

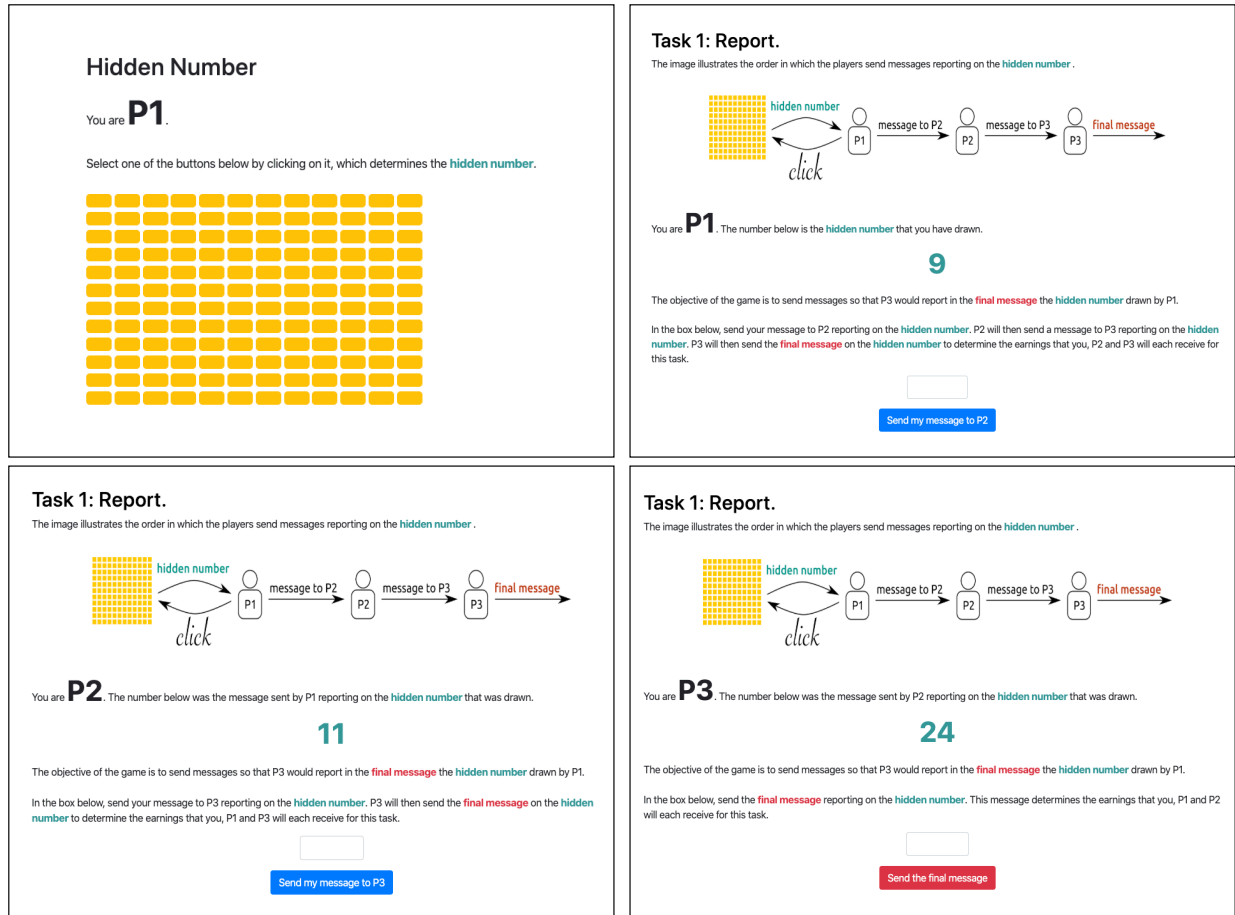


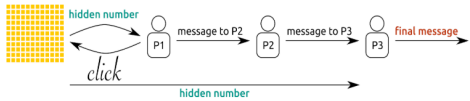
Figure F1. Decision screens in NO

Note: In NO, the *web of lies* game begins with (Upper-left) P1 choosing a card from a 12 × 12 grid by click on it, each card with an integer between 1 and 30, (Upper-right) P1 then reports a number to P2, (Lower-left) P2 observes the number sent by P1 and reports a number to P3, (Lower-right) P3 observes the number sent by P2 and sends a report to the experimenter.

Treatment EXO

In Figure F2 we display the decision screens for the last player (P3) in treatment EXO.

Task 1: Report.
The image illustrates the order in which the players send messages reporting on the **hidden number**.



You are **P3**. The number below was the message sent by P2 reporting on the **hidden number** that was drawn.

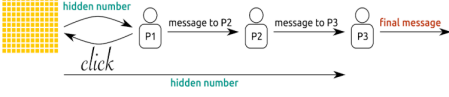
24

You have not been selected by chance to see **hidden number** that P1 drew.
The objective of the game is to send messages so that P3 would report in the **final message** the **hidden number** drawn by P1.

In the box below, send the **final message** reporting on the **hidden number**. This message determines the earnings that you, P1 and P2 will each receive for this task.

Send the final message

Task 1: Report.
The image illustrates the order in which the players send messages reporting on the **hidden number**.



You are **P3**. The number below was the message sent by P2 reporting on the **hidden number** that was drawn.

24

You have also been selected by chance to see **hidden number** that P1 drew, which was:

9

The objective of the game is to send messages so that P3 would report in the **final message** the **hidden number** drawn by P1.

In the box below, send the **final message** reporting on the **hidden number**. This message determines the earnings that you, P1 and P2 will each receive for this task.

Send the final message

Figure F2. Decision screens in EXO

Note: In EXO P1 and P2 make decisions with identical information as in NO. (Left) P3 has the same info in EXO as in NO, only observes the report from P2, with 20% chance. (Right) P3 observes both the report from P2 and the hidden number with 80% chance.

Treatment ENDO

In Figure F3 we display the decision screens for the last player (P3) in treatment ENDO.

Task 1: Report.
The image illustrates the order in which the players send messages reporting on the **hidden number**.

You are **P3**. The number below was the message sent by P2 reporting on the **hidden number** that was drawn.

24

You can see the **hidden number** P1 drew by clicking on the show button or proceed to send the final message by clicking on continue:

Task 1: Report.
The image illustrates the order in which the players send messages reporting on the **hidden number**.

You are **P3**. The number below was the message sent by P2 reporting on the **hidden number** that was drawn.

24

The objective of the game is to send messages so that P3 would report in the **final message** the **hidden number** drawn by P1.

In the box below, send the **final message** reporting on the **hidden number**. This message determines the earnings that you, P1 and P2 will each receive for this task.

Task 1: Report.
The image illustrates the order in which the players send messages reporting on the **hidden number**.

You are **P3**. The number below was the message sent by P2 reporting on the **hidden number** that was drawn.

24

The **hidden number** that P1 drew was:

9

The objective of the game is to send messages so that P3 would report in the **final message** the **hidden number** drawn by P1.

In the box below, send the **final message** reporting on the **hidden number**. This message determines the earnings that you, P1 and P2 will each receive for this task.

Figure F3. Decision screens in ENDO

Note: In ENDO P1 and P2 make decisions with identical information as in NO. (Upper-Left) P3 has the option to click on the “Show the hidden number” button to verify it or to click on the “Continue” button and avoid seeing the hidden number. (Upper-Right) If P3 avoids verification, he only observes the report from P2. (Bottom) If P3 decides to verify, he observes both the report from P2 and the hidden number.