Blue lies*

Gary Charness and Daniela Grieco

Abstract: *In the "blue lie game", subjects can decide to lie in favor of themselves and also of their ingroup. We find that blue lies are pervasive and detrimental for the society. They decrease in the presence of an ingroup norm of truth-telling, but increase when the same norm is exhibited by the outgroup. Stronger ingroup identity determines a higher sensitivity to the ingroup norm. Silence substantially reduces the number of blue lies, weakens the effects of norms, and increases social welfare. We structurally estimate a model that disentangles the possible drivers of blue lies and evaluates social welfare.*

Keywords: lying behavior; in-group bias; norms; identity; silence.

JEL codes: C91, D63, D71

* Corresponding author: Daniela Grieco, University of Milano, Department of Law C. Beccaria, via Festa del Perdono 7, 20122, Milano, Italy. Email: daniela.grieco@unimi.it Gary Charness, deceased on May 17, 2024, beloved friend, coauthor and mentor. We acknowledge the financial support from Fondazione Cariplo and the University of Milano. The study received ethics approval from the University of Milano Ethics Committee (Parere 69/2023). The experiment was pre-registered on December 10, 2023 in the AsPredicted database under trial title 'Blue lies' (#154688): https://aspredicted.org/wi2x6.pdf

The control treatment that tests the difference between blue lies and white lies was pre-registered on June 18, 2024 in the AsPredicted database under trial title 'Blue lies (cont.): comparison with white lies' (#179393): https://aspredicted.org/hu47n.pdf. An additional control treatment with no lies was pre-registered on September 11, 2024 in the AsPredicted database under trial title 'Blue lies (cont.): no information' (#189753): https://aspredicted.org/im95x.pdf

For valuable comments, we thank Isabel Brocas, Juan Carrillo, Yan Chen, David Dillenberger, Uri Gneezy, Ismael Rodriguez-Lara, the participants in the 2nd Science for Decision Making Conference in Hong Kong, the 2024 ESA World Meeting, and invited seminars at UCSB, USC, and Strasbourg University.

1. Introduction

The concept of "blue lies" refers to lies that are made in favor of the group to which a person belongs (the "ingroup"), and to the detriment of non-group members (the "outgroup"). Coined by the criminologist Carl Klockars with reference to the blue color of U.S. police uniforms, the term was meant to characterize a lie told in order to achieve compliance avoiding the use of force. However, the term later became generalized to all lies told by police, and then especially to the lies told by police officers to conceal questionable behavior on the part of a police force. By extension, the term came to mean any lie told to protect one's group or organization (Hornung, 2016; Borsellino, 2013) to the harm of others. Since adding this *ingroup-outgroup* dimension, blue lies differ from both white lies, which go in favor of others, and black lies, which go against others.[1] Blue lies exist because individuals identify with groups, but are not coincident with "collective" or "collaborative" lies (Leib et al., 2021): in fact, the decision to blue-lie is made at the *individual* level.

In this paper, we attempt to reproduce the blue-lie phenomenon in the lab, in order to investigate how detrimental blue lies are for the society, and how they can be discouraged. We introduce a new game where a subject (the sender) decides to send a true or false message trying to convince an opponent (the receiver) to choose between two options, as in Gneezy (2005). The possible lie, if the receiver believes to it, causes her to pick the option that gives the sender a relatively higher payoff. An unusual feature in our design is that the payoff is higher not only for her, *but also for her ingroup member*, with lower payoffs to the receiver *and to the other outgroup member*.[2] A control treatment with white lies – which, on the contrary, *do not go against* the outgroup – is run to disentangle the cost of *blue*-lying from the cost of lying *per se*.

The rich literature on ingroup bias and outgroup prejudice (e.g., among many others, Charness et al., 2007 and Chen and Li, 2009) shows that we are intensely social creatures but, especially when it comes to allocate resources, we are prone to divide ourselves into competitive groups. Early laboratory work by Tajfel et al. (1971) showed that simply categorizing participants into different groups based upon a trivial criterion (like flipping a coin or choosing a painter) engendered a positive bias toward one's group. Ingroup bias has been shown to be malleable (Charness and Chen, 2020). Individuals who identify strongly with their group have been found to display more prosocial or helping behavior towards ingroup members, together with greater antisocial behavior toward outgroup members (Hornstein, 1976; Ikeda et al., 2018; Dimant, 2024).

[1] See Erat and Gneezy (2012) for a taxonomy about white and black lies, and Levine and Schweitzer (2015) on the effects of prosocial lies.
[2] This feature was also used in the Charness et al., (2007) group-membership study.

Nonetheless, the causal link between identity and intergroup relations may also go in the opposite direction: competition and conflict among groups could reinforce ingroup's sense of identity (e.g., Bernard, 2012).

Thus, social identities cause people to flexibly "tune" their reactions to moral contexts (Van Bavel et al., 2022; Ellemers et al., 2019), making immoral behavior tolerable when directed to the outgroup. Although normally (at least in the past) lying is morally condemned, it may be justified when it serves pro-ingroup purposes (Barnes, 1994; Bok, 1978). For instance, lying about the outcome of a play in a team sport helps the team, and team members might consider such a behavior acceptable or even mandatory. When one knows the results of the play were actually negative but have been ruled to be positive, one faces an inner struggle between honesty and some form of loyalty. Likewise, citizens accept that intelligence agencies lie in the interests of national security, and laud own spies as heroes (Smith, 2017).

Lies and tendentious simplifications are widespread in politicians' speeches. Politicians often make misleading statements that are easy to unmask, but being exposed seems not to weaken the consensus of their supporters. On the contrary, their lies have the aim of remaking an ingroup's values and boosting the strength of identity for voters close to them on an ideological dimension (Besley and Coate, 1997). This happens especially in the case of important economic and social issues, where divisions between interest groups can easily lead to polarization and intergroup hostility (Fisher, 2000). Recent times have seen polarization and radicalization increasing dreadfully, exacerbated by the unregulated use of social media that tends to create "echo chambers" or "filter bubbles".[3]

Politicians ride this wave to gain voters by generating in them a sense of belonging. The expression "fake news" has become common since Trump's first electoral campaign (Farhall et al, 2019), where he "weaponized" the term to attack and discredit mainstream news media and political rivals (McNair, 2018, p. 91). As emphasized by Allcott and Gentzkov (2017), fake news is "distorted signals uncorrelated with the truth" (p. 212) that arises because news consumers enjoy partisan information that confirms their (deeply-held) priors (Mullainathan and Shleifer, 2005; Gentzkow and Shapiro, 2006; Chopra et al., 2023). Trump's lies only *strengthen* his support, with his endorsement among Republicans *increasing* every time a lie was exposed (Smith, 2017). This might stem from voters associating themselves with their party at such a visceral level that any

[3] These prevent people from learning information contradicting their pre-existing beliefs (Bail et al., 2018; Sunstein 2001 and 2007; Pariser, 2011), particularly in circumstances where relevant voices have been actively excluded and all outside sources are systematically distrusted (Nguyen, 2020). Social media also encourages people to share opinions in any domain, no matter the truth: anonymity and computer-mediated communication promote uninhibited expression (Siegel et al., 1986; Wu and Atkin, 2018), often deceitful, that may fuel conflict.

criticism to the party feels like a threat to the self, and triggers such a defensive psychological process (Theodoridis, 2017).

While the economic literature on lies and deception at the individual level is extremely rich (since the seminal work by Gneezy, 2005), to our knowledge there are no studies on the *blue lies game*. For a lie to be defined as "blue", and not simply as "collective" or "collaborative", group members need to *individually* lie in favor of the group. On the contrary, the literature on "collaborative dishonesty" (Lieb et al., 2021) focuses on how groups get collectively, simultaneously, or sequentially to the lying decision, which is a *joint* decision. Furthermore, we depart from Aksoy and Palma, 2019, and Cadsby et al., 2016 in having a game where lies recipients *do make decisions* on following or not the lies, determining the final payoffs. Using a die-roll task, Aksoy and Palma (2019) show that subjects in naturally-occurring groups cheat in favor of an ingroup member. Similarly, Cadsby et al (2016) find that people cheat on behalf of a member of their own ingroup at the expense of a non-member even when this choice does not affect their own monetary payoff. Differently from our study, these works address cheating behavior whose consequences affect two passive recipients (one from the ingroup, one from the outgroup) only. On the contrary, in our design the potential liar interacts with the receiver: *both* their decisions determine ingroup's and outgroup's payoffs, allowing to study how the liar weights the relative importance of her own payoff, the ingroup peer's payoff, the outgroup's payoff, and the costs of lying, and how the victim interprets and reacts to her messages.

Our game is used to study three aspects. The first is the role of social identity in affecting blue lies. We compare two types of groups: those assigned randomly, and those based on political orientation. Since identifying with the group is shown to enhance members' motivation to pursue group interests and goals, we investigate whether blue lies increase the stronger the identification with the group (which we elicit), and whether such identification is stronger when the partition depends on polarizing aspects such as politics (Dimant et al., 2023; Lane et al., 2024).

If partisan identity affects the frequency of blue lies, as we expect, our second objective is to understand why. A possible explanation may rely on norms: shared identity means expecting that ingroup members will share the same norms and that outgroup members will not (Choi, 2023). On the one hand, receiving information on the ingroup's lying decision might cause an update in the beliefs on what is the correct behavior to choose according to the group. On the other hand, obtaining information on the outgroup's lying decisions might generate a desire to differentiate from them, affirming one's own moral superiority. In this respect, field studies on collective lies regarding children (Fu et al., 200 and 2008; Sweet et al., 2010; Lau et al., 2012; Lu et al., 2016; Carl and Bussey, 2022) show that the frequency of lying depends on cultural upbringing. In Fu et al.

(2008), a sample of Chinese children had to create a team of four students who would compete in a chess contest in accordance with selection rules that induced them to under-report their skill. The proportion of subjects who lied in order to increment the chance for the team to win the contest increased with age. This suggests that as Chinese children become more exposed to their culture, their conception of lying aligns with the cultural norms prevailing in the society where they live. Euro-Canadian children, on the contrary, were more likely to morally condemn collective lies as age increased (Fu et al., 2007; Lau et al., 2012; Carl and Bussey, 2022; Sweet et al., 2010)[4]. The explanation seems to depend on cultural upbringing: while Western societies emphasize the importance of individual rights to information and freedom of choice, in collectivistic environments like the Chinese society the focus is on group cohesiveness: when a truthful statement harms group cohesiveness, truth-telling is no longer perceived as morally preferable and is discouraged, while lying tends to be condoned (Fu et al., 2007).

As a second contribution, we thus test whether increasing awareness of ingroup's (versus outgroup's) norms - by informing people of the ingroup and outgroup descriptive norms - may influence the frequency of lies. In particular, we truthfully report information on the behavior of previous participants with the goal to affect participants' beliefs of what is commonly done in this game (for applications in other related contexts see Hallsworth et al., 2017; Damgaard and Gravert, 2018; Allcott and Kessler, 2019; Bursztyn et al., 2020a; 2000b; Dimant et al., 2020; Dimant, 2023).

Group members are expected to conform to and enforce group norms (Van Bavel and Packer, 2021) and this may have beneficial consequences when exposed to a positive example as in our intervention. On the other hand, a good example can be detrimental if it comes from the outgroup (Berger and Heath, 2008). Looking at the interaction between norms and perceived identity allows us to understand the extent to which shared identities translate into shared values and behavior (e.g., Crandall et al., 2002; Murrar et al., 2020) and to test whether there is heterogeneity in reacting to ingroup or outgroup norms.

As a third contribution, we study a behavioral intervention in the form of a nudge: we change the set of available options (blue-lying versus telling the truth) by adding the possibility to remain silent, i.e., send no message. The rationale is to explore the effects of a simple modification of the choice set that comes at no cost and allows one to not behave in any partisan manner.[5] We

[4] Sweet et al (2010) test children who were asked to evaluate stories in which protagonists either lied or told the truth about their group's transgression, finding that Chinese children evaluated lies told to conceal a group's transgressions less favorably than did US children.

[5] To our knowledge, there is no previous investigation on the effect of *discouraging* communication in interactive intergroups, although silence has already been shown to be "golden" in a few cases of teamwork (Charness et al., 2020; Sio et al., 2018). Some evidence in line with our result is provided in Levine et al. (2018), where subjects are asked to

test whether saying explicitly that participants can abstain from sending a message might reduce intergroup antisocial behavior in the form of blue lies. The counterpart of this manipulation in the field corresponds to reminding or making salient the possibility of not expressing any opinion. On the contrary, algorithms governing social-media interactions are set to encourage comments and reactions, often fueling storms and hate speech in self-reinforcing spirals: the higher the number of people having the same opinion, the easier to convince others (Sznajd-Weron and Sznajd, 2000; de Arruda et al., 2020). Indeed, the increased presence of hate speech in one's environment works as “a descriptive norm” that makes outgroup denigration acceptable (Bilewicz and Soral, 2020, p. 3).

Our results show that although partisan identity is not *per se* associated with higher rates of blue lies, it nevertheless induces an increase in the compliance to ingroup norms: when informed that the majority of ingroup members do not tell blue lies, subjects tell blue lies at a significantly lower rate. However, when informed that the majority of *outgroup* members do not tell blue lies, subjects tell significantly *more* blue lies. Allowing silence dramatically reduces blue lies, weakens the effectiveness of norms, and increases social welfare. Textual analysis shows that blue lies are driven by monetary concerns, that are mentioned in 55% of the open comments that respondents write to explain their choice: subjects explicitly mention their ingroup and/or outgroup in 41% of the answers, while honesty concerns motivate the truth-telling and is present in 47% of the answers. Partisan concerns become more important the higher the level of partisan identity, and this happens at the expenses of honesty concerns.

The remainder of the paper proceed as follows. Section 2 describes the experimental design, hypotheses, and procedure. Section 3 presents the results, in the form of summary statistics, reduced-form estimates and textual analysis. Section 4 provides a theoretical model, the structural estimates of its parameters and a welfare analysis. Section 5 concludes.

2. The experiment

2.1 Experimental design

The experiment has a 2x2x3 design, with two pairs of conditions and three treatments, plus two controls. The conditions are “Random” vs. “Political Orientation”, differing in the way how groups are formed, and “NoSilence” vs. “Silence”, differing on the menu of available choices. Furthermore, we have three treatments: “NoNorm”, “IngroupNorm”, “OutgroupNorm”, differing in

rate acceptability and benevolence of omission (versus commission and honesty) in hypothetical conversations between an oncologist and a cancer patient: those who communicate information on average believe that omission is more ethical than telling a lie, whereas the targets of this information tend to believe the opposite.

the information given about other participants. A control treatment with white lies ("White Lies") is run for the "Random" vs. "Political Orientation", and the "NoSilence" vs. "Silence" conditions. A further control with no information transmission and thus no possibility to lie ("No Info) is used as benchmark. The details of the design are explained below, after a brief sketch of the general structure of the experiment. The survey is in the Supplementary Material.

Participants are assigned to two groups (Green and Yellow), both comprised of two people. One of the two members of the Green Group, selected at random, is asked to send a message to the Yellow Group. For the sake of clarity in describing the experiment, we call her "Green sender". The Green sender is informed that her message will reach one of the two members of the Yellow Group, selected at random. Again, for the sake of clarity we call her the "Yellow receiver". The Yellow receiver reads the message and makes a decision between options A and B without knowing the payoff of each option, having only the information contained in the message from the Green sender.

Option A gives \$2 to each member of the Yellow Group, and \$1 to each member of the Green Group. Option B gives \$2 to each member of the Green Group, and \$1 to each member of the Yellow Group. The Green sender decides between two messages: Message 1 ("Option A is more favorable for you than Option B", i.e., tells the truth) and Option 2 ("Option B is more favorable for you than Option A", i.e., lies). Finally, the Yellow receiver makes the decision between Option A and Option B. Since blue lies are by definition lies that benefit the ingroup at the expenses of the outgroup, for simplicity we set the payoffs to be symmetric across Green members. The same holds for the Yellow group members.

In the White Lies control treatment, Option A gives \$2 to each member of the Yellow Group, and \$1 to each member of the Green Group, as in case of blue lies. The difference is that *also Option B gives \$2 to the Yellow Group*, besides giving \$2 to the Green Group, as above. Thus, the Yellow group receives *\$2* no matter the selected option. The Green sender decides between two messages: Message 1 ("Option A is *equally* favorable for you than Option B", i.e., tells the truth) and Option 2 ("Option B is more favorable for you than Option A", i.e., lies). This time, lying has no detrimental effect for the Yellow Group (whose members get \$2 in any case), but goes in favor of the Green Group only. Following Erat and Gneezy (2012), this type of lies can be defined as "Pareto white lies".

In the No Info control treatment, Option A and Option B are the same as in the Blue Lies treatments. What differs is that the Green sender does not send any information (neither a message nor the fact that she prefers not to send any message) to the Yellow receiver, who decides totally blindly between Option A and Option B.

We now describe how this general structure is modified by specific conditions or treatments.

"Random" versus "Political Orientation": This condition pertains to the manner in which the two groups are formed. In the Random condition participants are drawn from the general population on Prolific without any pre-screening, and they are randomly assigned to the Green or to the Yellow group. In the Political Orientation condition, participants are drawn from among those who previously pre-screened as Democrat or Republican[6], such that the Green group is comprised of two Republicans (Democrats) who are aware that their group is made of Republicans (Democrats) and that they will interact with a group made of two Democrats (Republicans). To measure the strength of ingroup identity, we subjects assign a score between 1-10 to capture how much they identified with their ingroup (10 is the highest level of identification).[7] The question is asked of participants in both conditions. Furthermore, to assess the possible effect of making ingroup identity salient by asking the question, we randomize the position of this question, with half subjects receiving the question *before* the blue-lying decision, and half subjects receiving it *after* the blue-lying decision.

"NoSilence" vs. "Silence": These conditions differ in the set of choices presented to the Green Sender. In the NoSilence condition, the Green sender has two possibilities only: Option A (which corresponds to sending a message that tells the truth) or Option B (sending a message that contains a lie). In the Silence condition, the Green sender has *three* possibilities: sending Option A (sending a message that tells the truth), Option B (sending a message that contains a lie), or not sending any message (being "silent").

"NoNorm" vs. "IngroupNorm" vs. "OutgroupNorm": These treatments differ among each other on the information that is given to the Green sender. While in the NoNorm treatment she has no information of what other participants have chosen, in the IngroupNorm she is informed about what the majority of other Green senders – in a sample of respondents who previously took part in the same study – did.

When groups are assigned randomly, as in the Random condition, there is no defined distinction between Ingroup and Outgroup, and we may assume that all the other Green subjects are part of the same ingroup. On the contrary, when groups are assigned depending on political orientation, the Green sender could be stated to be either a Republican or a Democrat. In this case,

[6] Participants were reminded of their pre-screening choice through a comprehension check that reads as follows: "You have been screened on the basis of your political orientation as declared when subscribing to Prolific. Together with another respondent who identified her/himself as 'Republican" ("Democrat"), like you, you have been assigned to a group of two people called "Green". Instead, two other respondents who identified as "Democrat" ("Republican") are assigned to a group called "Yellow". Is this clear? To check your attention level, we ask you to indicate the group you have been assigned to."

[7] The question was phrased as follows: "How much do you identify with the Green group, which you have been assigned to at random/which is composed by Democrats/Republicans?". Respondents could move the slider between 0 ("Not at all") and 10 ("A great deal").

the IngroupNorm treatment gives to a Republican (Democrat) Green sender information on what the majority of other Republican (Democrat) Green senders did. The OutgroupNorm treatment, in contrast, gives information on what a Green sender of opposite political orientation did.

In this paper, we are mainly interested in the sending of blue lies rather than the reaction to possible blue lie, so that treatments and conditions only affect the instructions for the Green sender's (the Yellow receiver always reads the same instructions, with the exception of the NoInfo treatment where the Yellow does not receive any type of information from the Green).[8] What may vary, depending on being in the "Silence" vs. "NoSilence" condition and on the choice made by the Green sender, is receiving a message from the Green sender or not, and the content of the message. If the Green senders decides not to send any message, the Yellow receiver is informed that, notwithstanding the possibility to send a message, the Green sender decided not to do it.

2.2 Hypotheses

Hypothesis 1 ("group identity"). Partisan identity increases the frequency of blue lies.

One characteristic of blue lies, as opposed to non-group lies (the ones commonly known as "black lies"), is that the potential benefits of lying affect not only the liar herself, but also the liar's ingroup. Thus, when choosing between lying or not, the individual must consider that not lying diminishes not only her payoff, but also her peers' ones. A higher degree of "attachment" or identification with the ingroup is then expected to increase the rate of blue lies. But what determines identification with the ingroup or "group identity"?

On the one hand, when rewards are structured such that all ingroup members share equally the fruits of labor or luck, this can be seen as a "common fate situation" (Hornsey et al, 2003). Common fate can be defined as "a coincidence of outcomes among two or more persons that arises because they have been subjected to the same external forces or decision rules" (Brewer, 2000, p. 118). The power of shared rewards to promote intergroup cohesion is grounded on the fact that "it does not rely on higher order principles of tolerance or altruism; rather, it trades on people's self-interest" (Hornsey et al, 2003, p. 275). Thus, a shared self-interest might be enough to generate a sense of group cohesion and foster blue lies.[9]

[8] These are the instructions given to the Yellow sender in the Random conditions: "A member of the Yellow Group – selected at random between you and your peer – will make a decision that determines the payments for you and the other member of the Yellow group, and for the two members of the Green group as well. You have been selected. You are the member of the Yellow group who has to choose between two options (Option A and Option B) without knowing which payments are associated with each option. The only information you will have is information we will send to you in a message. This is the message sent to you by one member of the Green group, who knows the payoffs associated to the two options: "Option A (B) is more convenient for you and your group"."

[9] Shogren (1989) investigates a similar situation, finding that loyalty trumps fairness.

On the other hand, there are divisions that can generate a strong sense of identity *per se*. Among these, political orientation has been shown to be a highly-polarizing issue (e.g., Dimant, 2024) that can generate a strong sense of identity across individuals sharing the same political ideology. For this reason, we compare blue-lying behavior across groups involving political orientation (U.S. Democrats versus Republicans) and groups formed at random. In both cases, we elicit the self-reported degree of identification with the ingroup.

Hypothesis 2 (norms). People tend to mimic the perceived ingroup norms and do the opposite of the perceived outgroup norms.

Norms have been widely shown to shape individuals' beliefs and behavior. They summarize the observed prevalent behavior (or the behavior that is considered appropriate). Therefore, they are a product of group's values and are likely to be more influential the more the individual identifies with the group. Our design compares the effects of providing information on the ingroup's and on the outgroup's descriptive norms by telling what the majority of group members did while taking part in previous sessions of the same study. The literature suggests that subjects tend to conform to the behavior of the ingroup, and thus to comply to the ingroup norm, with this effect stronger the higher the identification with the ingroup: individuals who identify with a given group seek to resemble typical group members, especially along salient dimensions (Shayo, 2020). On the other hand, we expect subjects' behavior to be uncorrelated with the behavior of the outgroup, or even be negatively correlated. Our experimental design allows us to compare how the rate of blue lies varies when the same norm is attributed to the ingroup or to the outgroup. To test the role of ingroup identification in fostering the relevance of norms, we exploit the comparison between random-based and political-orientation-based group assignment.

Hypothesis 3 (silence). The frequency of blue lies decreases when people can choose silence.

The possibility of silence represents, in our set up, a *tertium quid* on top of lying and telling the truth. Thus, the first immediate consequence of adding the option to be silent is enlarging the set of options available to the decision-maker: having three options instead of only two can "mechanically" decrease the rates in the original two categories. Furthermore, in case of blue lies, the option to remain silent can translate to not forcing one to take a side when one does not really want to do so. Indeed, subjects could in principle prefer being silent. In the field, silence is often the rule (consider Russia under Putin or omertous behavior in mafia-like environments, for example). Inducing people to speak when seeking their opinion might, on the one hand, encourage less engaged individuals to communicate how they feel and what they think, but also stimulate comments that they had originally considered unnecessary. It also seems true that individuals tend to comment more frequently when they have very strong opinions. So, asking for the opinions of all people involved could lead to collecting more moderate views, perhaps producing a more representative average outcome. Our design allows us to isolate the effect of adding the silence option since it randomizes the presence of this option across all the treatments.

A somewhat similar interpretation sees the option of being silent as a "compromise choice", i.e., a compromise between the two extreme alternatives of telling the truth or the lie. The compromise effect occurs when an alternative becomes more popular for the only fact of representing a middle option in the choice set (Simonson, 1989). The compromise effect helps explain several interesting facts, such as the finding that brands gain share when they become the intermediate rather than extreme option (Kivettz et al., 2004). Simonson and Tversky (1992) explain the compromise effect by the theory of "loss aversion" or "extreme aversion": in the presence of the chance of selecting a compromise option, "the disadvantages of the extreme options loom larger than the advantages" (Boldt and Arora, 2017, p. 438). Thus, in our context, having the chance of being silent could make the cost of the blue lie higher, reducing blue-lying behavior.

Hypothesis 4 (silence and norms). The possibility of silence decreases the influence of norms.

As emphasized in Hypothesis 3, adding the option of being silent constitutes an enlargement of the range of options. To our knowledge, there is no current evidence about the consequences of delivering a norm in a situation where the agents may either follow the norm, behave in contrast to the norm, or have a third possibility consisting of abstaining from both the previous actions. Thus, there is no clear *a priori* indication on how norms and silence would interact. One effect that might be expected is that the silence option may "cloud" the strength of a norm, since it provides a further possibility that is not the one directly contemplated in the norm.

Hypothesis 5 (blue versus white lies). *White lies are more frequent that blue lies.*

By definition, white lies have no detrimental effect on the outgroup. As emphasized in Erat and Gneezy (2012), Pareto white lies, like ours, entail only the pure cost of lying, which a person faces if she considers lying as a bad act *per se*, since its outcome has no consequences on other people. If individuals do face a moral cost for their lies to be detrimental on others, we can expect the cost of blue-lying to be higher than the cost of white-lying (that is made only of the cost of lying *per se*), and thus the choice of blue-lying to be less frequent.

2.3. Experimental procedure

The experiment was conducted between 13 December 2023 and 16 September 2024. A total of 9672 U.S. citizens, fully proficient in speaking English, were recruited on Prolific (see Palan and Schitter, 2018 for a discussion of the platform and subjects' recruitment) to have a sample gender-balanced sample. We ran 48 sessions in total, 24 for Green subjects and 24 for Yellow subjects, with 32 sessions regarding blue lies, 12 for the control treatment involving white lies and 4 for the control treatment with no information. Since some sessions required a pre-screening based on political orientation, we implemented the experiment sequentially, first recruiting the Green subjects (two in each group) to collect information on the message they decide to send to the Yellow subjects, then recruiting the Yellow subjects, transmitting to each Yellow group the message of the Green sender randomly associated to them. The decision by the Yellow receiver was used to distribute payoffs, which could be either \$2 or \$1. The two Green subjects in each group were paired at random; similarly for the two Yellow subjects in each group. The Green subjects who had to choose a message were also asked to motivate their choice after they made it: they received an additional \$0.10 for providing a motivation that was "sufficiently accurate". The experiment lasted on average 167 seconds for the Green subjects and 116 seconds for the Yellow subjects[10]. Tables A1 and A2 in the Appendix provide an overview of the sessions for Green and Yellow subjects, respectively.

3. Experimental results.

3.1. Descriptive statistics

We organize our experimental results according to the hypotheses formulated in section 2.2. With the exception of the last hypothesis, that compares blue and white lies, the following results refer to blue lies only.

Ingroup identity. Table 1 displays self-reported levels of identification with the ingroup in the treatments aimed at testing the role of identity, namely the Random versus Political Orientation ("PO") conditions, distinguishing between sessions where the question on identification with the group was asked before ("PRE") or after ("POST") the blue-lying decision. For the PO condition, we distinguish between Democrats and Republicans.

The Table shows that self-reported identification levels, elicited before the blue-lie decision, increase significantly when groups are formed depending on political orientation, either if we compare randomly-formed groups to political orientation-based groups without distinguishing between Democrats and Republicans (Z = -8.723, $p < 0.001$, two-sample Wilcoxon rank-sum test),

[10] This average considers only sessions with Blue or White lies. In the No Info treatment, where there was no message sent or received, the experimental sessions lasted on average 66 seconds.

or if we compare them separately (Random versus Democrats: Z = -7.892, $p < 0.001$, two-sample Wilcoxon rank-sum test; Random versus Republicans: Z = -7.176, $p < 0.001$, two-sample Wilcoxon rank-sum test). This also holds if we elicit the question *after* the blue-lie decision, either if we compare randomly-formed groups to political orientation-based groups without distinguishing between Democrats and Republicans (Z = -9.167, $p < 0.001$, two-sample Wilcoxon rank-sum test), or if we compare them separately (Random versus Democrats: Z = -7.282, $p < 0.001$, two-sample Wilcoxon rank-sum test; Random versus Republicans: Z = -8.591, $p < 0.001$, two-sample Wilcoxon rank-sum test). Eliciting before or after the blue-lie decision makes little difference: Pre versus Post with Random groups: Z = 0.750, $p = 0.453$, two-sample Wilcoxon rank-sum test; Pre versus Post with political orientation-based groups: Z = -0.485, $p = 0.678$, two-sample Wilcoxon rank-sum test.

Table 1. Self-reported levels of identification with the ingroup

Condition	Order	Average	Std. Err	Obs.
Random	PRE	5.010	0.203	203
PO Dem	PRE	7.228	0.162	202
PO Rep	PRE	7.015	0.160	197
Random	POST	4.820	0.206	200
PO Dem	POST	6.861	0.149	201
PO Rep	POST	7.307	0.161	199

The table reports subjects' answers to the question "How much do you identify with the Green group, which you have been assigned to at random/which is composed by Democrats/Republicans?", ranging from 0 ("Not at all") to 10 ("A great deal"). Subjects in the White lies control and in the No lies control treatments are not included. Random refers to groups' random assignment, PO to groups' assignment based on Political Orientation. Dem refers to subjects who pre-screened as Democrats, while Rep refers to subjects who pre-screened as Republicans. PRE refers to sessions where the question on group identity was asked before the blue-lying choice, while POST refers to sessions where this question was asked after the blue-lying choice. Note that the number of observations roughly doubles the ones in Table 2 because both Green subjects in each group answered to this questions, whereas only one in each pair took the blue-lying decision.

We can thus conclude that assigning groups at random versus assigning groups depending on a pre-screening based on political orientation is effective in engendering different levels of identification with the group: partisanship increases group identity significantly. Given this result, we turn to test the effect of ingroup identity on blue lies.

Table 2 summarizes the frequency of blue lies in the conditions aimed at testing the role of partisanship, as above. For now, we only focus on subjects who did not have the option to remain silent (thus those in the NoSilence condition). Table 2 provides an overview of the rates of blue lies across conditions varying how groups are formed. In the Political Orientation condition, the rate of blue lies is not significantly different from that of groups formed at random (Z= -0.938, $p = 0.348$, two-sample Wilcoxon rank-sum test). In this respect, there is no difference between Democrats and

Republicans (Z = 0.415, *p* = 0.678, two-sample Wilcoxon rank-sum test). Eliciting before or after the blue-lie decision makes no difference in groups based on political orientation: Pre versus Post with Democrats: Z = -0.307, *p* = 0.759, two-sample Wilcoxon rank-sum test; Pre versus Post with Republicans: Z = -0.984, *p* = 0.525, two-sample Wilcoxon rank-sum test), whereas we observe a slightly significant difference when comparing groups assigned at random: Pre versus Post with Random groups: Z = -1.996, *p* = 0.046, two-sample Wilcoxon rank-sum test.

Table 2. Frequency of blue lies and group identity

Condition	Order	Average	Std. Err	Obs.
Random	PRE	0.400	0.074	45
PO Dem	PRE	0.566	0.069	53
PO Rep	PRE	0.547	0.069	53
Random	POST	0.460	0.071	50
PO Dem	POST	0.380	0.069	50
PO Rep	POST	0.455	0.076	44

The table reports subjects' blue lies choices. Random refers to groups' random assignment, PO to groups' assignment based on Political Orientation. Dem refers to subjects who pre-screened as Democrats, while Rep refers to subjects who pre-screened as Republicans. PRE refers to sessions where the question on group identity was asked before the blue-lying choice, while POST refers to sessions where the question on group identity was asked after the blue-lying choice. Note that the number of observations is roughly one quarter of the ones in Table 1a because only one subject in each pair of Greens took the blue-lying decision, while both answer to the question on group identification, and because we consider only subjects in the NoSilence condition, which was administered to half of the Green subjects.

We summarize the main findings about partisanship, ingroup identity and blue lies in Result 1:

Result 1. Partisanship does not increase blue lies per se*: groups based on political orientation do not blue-lie significantly more than groups assigned at random.*

Ingroup versus outgroup norms. Table 3 summarizes the effect of providing a descriptive norm of ingroup's versus outgroup's behavior in the following form: "the majority of other Green Senders – in a sample of respondents who previously took part in the same study – selected Message A", where Message A corresponds to telling the truth. The information is not deceitful because, as shown in Table 2, there is at least one subsample of respondents in each treatment whose majority decides not to lie (this is always the case of the "PRE" condition, for instance).

Concerning the PO condition, the ingroup norm refers to the behavior of the majority of subjects sharing the same political orientation, whereas the outgroup norm is the behavior of the majority of subjects with opposite political orientation. In case of the Random treatment, the norm refers to the behavior of the majority of subjects assigned at random to the Green group, thus we do not distinguish between ingroup and outgroup norms and only speak of "norm"

Table 3. Frequency of blue lies and norms

Treatment	Identity	Average	Std. Err	Obs.
No norm	PO	0.490	0.035	200
Ingroup norm	PO	0.267	0.032	195
Outgroup norm	PO	0.549	0.036	193
No norm	Random	0.432	0.051	95
Norm	Random	0.292	0.043	113

The table reports subjects' blue lies choices. "No norm" refers to sessions where subjects do not receive any information on other subjects' previous blue-lying decisions. "Ingroup norm" refers to sessions where subjects receive information on ingroup (same political orientation) subjects' previous blue-lying decisions. "Outgroup norm" refers to sessions where subjects receive information on outgroup (characterized by opposite political orientation) subjects' previous blue-lying decisions. "Norm" refers to sessions in the Random condition where subjects receive information on the previous blue-lying decisions of other participants in the Random condition. PO to groups' assignment based on Political Orientation, Random refers to groups' random assignment.

Table 3 shows that conveying a norm of truthful behavior is highly effective in decreasing the rate of blue lies when the norm regards the ingroup (from 0.490 to 0.267: Z = -4.567, $p < 0.001$, two-sample Wilcoxon rank-sum test). We observe a reduction in blue lies also when there is no ingroup-outgroup dichotomy since the group assignment is random, but the effect is smaller and the significance weaker (from 0.432 to 0.292: Z = -2.089, $p = 0.037$, two-sample Wilcoxon rank-sum test). However, where the norm of truthful behavior is about the outgroup, we observe the opposite effect: the average rate of blue lies increases from 0.490 to 0.549. This difference is not significant by this test, but some regressions below do indicate statistical significance. In any event, the difference-in-difference across these cases is easily significant (Z = 5.801, $p < 0.001$, two-sample Wilcoxon rank-sum test).

We summarize the main findings about norms and blue lies in Result 2 below.

Result 2. Subjects conform to the ingroup norm, but they to not react to the outgroup norm.

Silence. Table 4 summarizes the effect of including the option of being silent (in addition to the options of telling either a blue lie or the truth) in the aforementioned conditions and treatments.

Table 4 shows that adding the option of remaining silent is always highly effective in decreasing the rate of blue lies. We observe a significant reduction in blue lies both when groups are assigned at random (from 0.432 to 0.288: Z = 2.100, $p = 0.036$, two-sample Wilcoxon rank-sum test), and when groups are based on political orientation (from 0.490 to 0.306: Z = 3.724, $p < 0.001$, two-sample Wilcoxon rank-sum test). When subjects receive the ingroup norm of truthful behavior, silence further reduces the average rate of blue lies.

Table 4. Frequency of blue lies and silence

Condition/treatment	Silence availability	Average	Std. Err	Obs.
Random	No silence	0.432	0.051	95
PO	No silence	0.490	0.035	200
Norm	No silence	0.292	0.043	113
Ingroup norm	No silence	0.267	0.032	195
Outgroup norm	No silence	0.549	0.036	193
Random	Silence	0.288	0.045	104
PO	Silence	0.306	0.033	193
Norm	Silence	0.279	0.049	86
Ingroup norm	Silence	0.217	0.029	207
Outgroup norm	Silence	0.383	0.034	201

The table reports subjects' blue lies choices. Random refers to groups' random assignment, PO to groups' assignment based on Political Orientation. "Norm" refers to sessions in the Random condition where subjects receive information on the previous blue-lying decisions of other participants in the Random condition. "Ingroup norm" refers to sessions where subjects receive information on ingroup (same political orientation) subjects' previous blue-lying decisions. "Outgroup norm" refers to sessions where subjects receive information on outgroup (opposite political orientation) subjects' previous blue-lying decisions. "No silence" corresponds to the condition when subjects can decide only between blue-lying or telling the truth, whereas "Silence" corresponds to the condition when subjects can blue-lie, tell the truth or remain silent.

However, the effect is not significant, perhaps because the ingroup norm has already caused a large reduction of lies. In any event, the average rate of blue lies decreases from 0.259 to 0.217 when silence is an option (Z = 0.978, $p = 0.328$, two-sample Wilcoxon rank-sum test). When the norm regards groups assigned at random, we still observe a decrease in blue-lying due to silence, although non-significant, from 0.292 to 0.279 (Z = 1.173, $p = 0.241$, two-sample Wilcoxon rank-sum test). Finally, when the norm is about the outgroup, silence does significantly reduce the rate of blue lies from 0.549 to 0.383 (Z = 3.301, $p = 0.001$, two-sample Wilcoxon rank-sum test). In the aggregate, the frequency of blue lies without a silence option is 0.413 and the frequency with a silence option is 0.297 (Z = 8.421, $p < 0.001$, two-sample Wilcoxon rank-sum test); providing a silence option thus reduces lies by 28.1%.

We summarize the main findings about silence, norms and blue lies in Result 3 below.

Result 3. Allowing for remaining silent significantly decreases the rate of blue lies and weakens the effects of ingroup norms.

It is important to note that the decrease in blue lies observed when allowing for silence comes together with a decrease in truth-telling: in aggregate, the frequency of truth-telling drops with a silence option from 0.587 to 0.375 (Z = 4.824, $p < 0.001$, two-sample Wilcoxon rank-sum test), reducing truth-telling by 36.1%. This decrease is observed not only in aggregate, but in all treatments. This finding corroborates the hypothesis that silence is a compromise choice between two costly options: telling the truth implies losing money in favor of the outgroup, while blue-lying has a moral cost, as we will discuss more in depth with the model presented in section 4.

An environment that promotes silence may, on one side, prevent the diffusion of lies, defamation and fake news. On the other side, silence has its own costs. In areas where mafia phenomena are widespread, the “conspiracy of silence” or "omertà” is a code that requires one to stay silent to avoid taking any responsibility. For Mafia members the “best word is the one you do not pronounce” and silence is a form of communication itself: “even what is unsaid reveals something” (Mannino et al., 2015). From the standpoint of social welfare, the question is whether the benefit from reducing blue lies is worth the cost of people who would have told the truth instead staying silent. This depends not only on the relative number of changes, but also on the consequences of silence and lies for the society. In a sense, this reflects omission versus commission, as in the classic trolley-car problem. Consider the current political environment and social media; are the lies themselves a bigger factor than the people who stay silent in their presence? Our welfare analysis below tries to address this (and other) questions.

3.2. Regression results

Table 5 confirms the results of non-parametric tests and also helps understanding the role of partisanship and group identity, and their interactions with group norms.

Column 1 shows that providing information on a prevailing truth-telling norm in the ingroup reduces blue-lying, whereas the same information about the outgroup *increases* blue-lying. Being allowed to remain silent significantly reduces the probability of blue lies. Being assigned to the group at random or according to political orientation plays no role. Column 2 shows that these results hold when controlling for the self-reported level of group identification, order effects and specific political leaning (Democratic versus Republican).

Column 3 introduces the interacted variable between group identification and ingroup norm, showing that the ingroup norm is no longer significant *per se*, but is only so for those subjects who highly identify with the group. Therefore, group identity plays a more subtle and complex role than the one suggested by the non-parametric tests conducted in section 3.1: belonging to a group of people sharing the same political orientation does not increase the likelihood of blue lies more *per se*, but it increases the extent to which a subject identifies with her own ingroup, and this shapes the sensitivity to the ingroup norm.

Column 4 shows the results of a similar exercise but related to the outgroup norm: indeed, it introduces the interacted variable between group identification and outgroup norm. Nonetheless, we observe no effect of group identity when the outgroup norm is provided. When all the interactions are considered together (column 5), we observe that a stronger group identity increases the rate of blue lies, notwithstanding the higher responsiveness to the ingroup norm (of not lying) of those who identify strongly with the ingroup, as noted above. Finally, the results hold also when correcting for Multiple Hypotheses Testing (column 6).

Table 5: Determinants of blue lies: identity and norms

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	treatments	identity	id&ingroup norm	id&outgroup norm	all	MHT
pol_orient	-0.006	-0.017	-0.014	-0.017	-0.026	-0.177*
	[0.081]	[0.092]	[0.093]	[0.093]	[0.093]	[0.095]
norm_own	-0.376***	-0.375***	0.005	-0.375***	0.077	0.157
	[0.077]	[0.077]	[0.187]	[0.077]	[0.201]	[0.209]
norm_others	0.211**	0.211**	0.201**	0.212	0.441*	0.974***
	[0.087]	[0.087]	[0.087]	[0.248]	[0.266]	[0.263]
silence	-0.326***	-0.326***	-0.320***	-0.326***	-0.320***	-0.365***
	[0.066]	[0.066]	[0.066]	[0.066]	[0.066]	[0.067]
ident_level		0.004	0.027	0.004	0.038*	0.054***
		[0.013]	[0.017]	[0.015]	[0.021]	[0.021]
pre_id		-0.018	-0.015	-0.018	-0.017	-0.050
		[0.066]	[0.066]	[0.066]	[0.066]	[0.066]
dem		0.010	0.009	0.010	0.010	0.041
		[0.076]	[0.076]	[0.076]	[0.076]	[0.076]
id_own_norm			-0.058**		-0.069**	-0.080***
			[0.026]		[0.029]	[0.029]
id_others_norm				-0.000	-0.034	-0.101***
				[0.033]	[0.036]	[0.035]
Constant	-0.129	-0.143	-0.301**	-0.143	-0.367**	-0.373**
	[0.081]	[0.114]	[0.135]	[0.119]	[0.152]	[0.160]
Observations	1,585	1,585	1,585	1,585	1,585	1,577

Probit (standard errors in parentheses). The dependent variable is a dummy variable assuming value equal to 1 when the subject lies, and 0 when she tells the truth. pol_orient is a dummy variable assuming value equal to 1 when the groups are assigned depending on a prescreening based on political orientation, and 0 when group assignment occurs at random. norm_own is a dummy variable assuming value equal to 1 when the subject receives information on the descriptive ingroup norm, and 0 when she receives no information. norm_others is a dummy variable assuming value equal to 1 when the subject receives information on the descriptive outgroup norm, and 0 when she receives no information. silence is a dummy variable assuming value equal to 1 when the subject has the option of being silent on top of lying and telling the truth, and 0 when she has no possibility of being silent. ident_level is a variable ranging between 1 and 10 that captures how much a subject identifies with the group she is assigned to. pre-id is a dummy variable assuming value equal to 1 when the question about the level of group identification is asked before the lying decision, and 0 when it is asked after the lying decision. dem is a dummy variable that assumes value equal to 1 when the subject has prescreened herself as "Democrat", and 0 if not. id_own_norm is the interacted variable between ident_level and norm_own. id_others_norm is the interacted variable between ident_level and norm_others.
*** significant at 1%; ** significant at 5%; * significant at 10%.

Result 4. Partisanship does not affect the likelihood of blue-lying per se, *but increases the extent to which a subject identifies with her own ingroup. This induces a higher sensitivity to the ingroup norm.*

Table 6 focuses on the interaction between silence and norms. The regressions show that adding the option of remaining silent significantly decreases the rate of blue lies, no matter the specification.

Table 6: Determinants of blue lies: silence and norms

VARIABLES	(1) identity	(2) id&silence	(3) silence&ingroup norm	(4) silence&outgroup norm	(5) all	(6) MHT
pol_orient	-0.017	-0.015	-0.024	-0.017	-0.022	0.053
	[0.092]	[0.093]	[0.093]	[0.092]	[0.093]	[0.096]
norm_own	-0.375***	-0.379***	-0.530***	-0.375***	-0.542***	-0.627***
	[0.077]	[0.077]	[0.101]	[0.077]	[0.106]	[0.110]
norm_others	0.211**	0.211**	0.217**	0.275**	0.207*	0.045
	[0.087]	[0.087]	[0.087]	[0.114]	[0.119]	[0.120]
silence	-0.326***	-0.516***	-0.440***	-0.292***	-0.661***	-0.901***
	[0.066]	[0.184]	[0.081]	[0.077]	[0.201]	[0.213]
ident_level	0.004	-0.009	0.003	0.004	-0.011	-0.006
	[0.013]	[0.018]	[0.013]	[0.013]	[0.018]	[0.018]
pre_id	-0.018	-0.018	-0.014	-0.017	-0.014	0.025
	[0.066]	[0.066]	[0.066]	[0.066]	[0.066]	[0.068]
dem	0.010	0.007	0.008	0.009	0.005	0.065
	[0.076]	[0.076]	[0.076]	[0.076]	[0.076]	[0.078]
id_silence		0.028			0.031	0.041
		[0.026]			[0.026]	[0.027]
sil_own_norm			0.328**		0.345**	0.571***
			[0.138]		[0.154]	[0.161]
sil_others_norm				-0.128	0.019	0.249
				[0.149]	[0.166]	[0.169]
Constant	-0.143	-0.056	-0.077	-0.158	0.025	-0.074
	[0.114]	[0.138]	[0.117]	[0.115]	[0.145]	[0.149]
Observations	1,585	1,585	1,585	1,585	1,585	1,532

Probit (standard errors in parentheses). The dependent variable is a dummy variable assuming value equal to 1 when the subject lies, and 0 when she tells the truth. pol_orient is a dummy variable assuming value equal to 1 when the groups are assigned depending on a prescreening based on political orientation, and 0 when group assignment occurs at random. norm_own is a dummy variable assuming value equal to 1 when the subject receives information on the descriptive ingroup norm, and 0 when she receives no information. norm_others is a dummy variable assuming value equal to 1 when the subject receives information on the descriptive outgroup norm, and 0 when she receives no information. silence is a dummy variable assuming value equal to 1 when the subject has the option of being silent on top of lying and telling the truth, and 0 when she has no possibility of being silent. ident_level is a variable ranging between 1 and 10 that captures how much a subjects identifies with the group she is assigned to. pre-id is a dummy variable assuming value equal to 1 when the question about the level of group identification is asked before the lying decision, and 0 when it is asked after the lying decision. dem is a dummy variable that assumes value equal to 1 when the subject has prescreened herself as "Democrat", and 0 if not. id_silence is the interacted variable between ident_level and silence. sil_own_norm is the interacted variable between silence and norm_own. sil_others_norm is the interacted variable between silence and others_norm.
*** significant at 1%; ** significant at 5%; * significant at 10%.

Column 2 shows that being allowed to remain silent significantly reduces blue-lying behavior no matter the degree of group identity, since the interacted variable between group identification and silence plays no role. Nonetheless, the effect of silence does interact with ingroup norms (column 3): silence "obfuscates" the effectiveness of ingroup norms, although this might be due to the fact that silence already reduces blue lies and ingroup truth-telling norms have less room to be effective. Silence has no differential effect depending on outgroup norms (column 4). Our results hold when considering together (in column 5) all the interactions analyzed separately in columns 2, 3 and 4, and when we control for Multiple Hypotheses testing (column 6), with the exception of the role of the outgroup norm, that – already only weakly significant - disappears.

3.3. Comparison between blue and white lies.

Table A3 and A4 in the Appendix summarize the frequency of white lies across conditions varying how groups are formed, and accounting for the possibility to remain silent. The frequency of white lies is significantly higher than the frequency of blue lies[11]: 0.528 versus 0.385 (Z= -4.927, $p < 0.001$, two-sample Wilcoxon rank-sum test). This holds both when group assignment occurs at random (0.549 versus 0.357, Z= -3.823, $p < 0.001$, two-sample Wilcoxon rank-sum test) and when is based on political orientation (0.518 versus 0.399, Z= -3.327, $p < 0.001$, two-sample Wilcoxon rank-sum test). When silence is available, and no matter the type of group assignment, the frequency of white lies is still significantly higher than the frequency of blue lies: 0.464 versus 0.300 (Z= -4.136, $p < 0.001$, two-sample Wilcoxon rank-sum test). Table A5 reports a set of regressions confirming these findings.

Result 5. The frequency of white lies is significantly higher than the frequency of blue lies.

Adding the option of remaining silent is effective in decreasing also the rate of white lies, but the difference is significant only when groups are based on political orientation (from 0.590 to 0.447: Z = 2.827, p = 0.005, two-sample Wilcoxon rank-sum test), while it is not when groups are assigned at random (from 0.611 to 0.495: Z = 1.617, p = 0.106, two-sample Wilcoxon rank-sum test).

3.4. Textual analysis

To better understand the previous results, we run a textual analysis of the content of subjects' written motivations that accompanied the choice between blue-lying, telling the truth and, when available, remaining silent. We conducted a topic modeling analysis on the full text of the comments to identify the major arguments the respondents brought when motivating their decision. The main keywords in each are different enough to identify three themes. One argument stresses the

[11] The comparison is carried out restricting to the "NoNorm" treatment which is available for both types of lies.

importance given to monetary concerns ("Monetary"). A second topic appeals to the moral duty of being honest ("Honesty"). The third theme refers to concerns related to the group dimension of decision and its consequences ("Ingroup").

The frequency of these themes is summarized in Figure 1.

Figure 1. Percentage of different motivations for choice made.

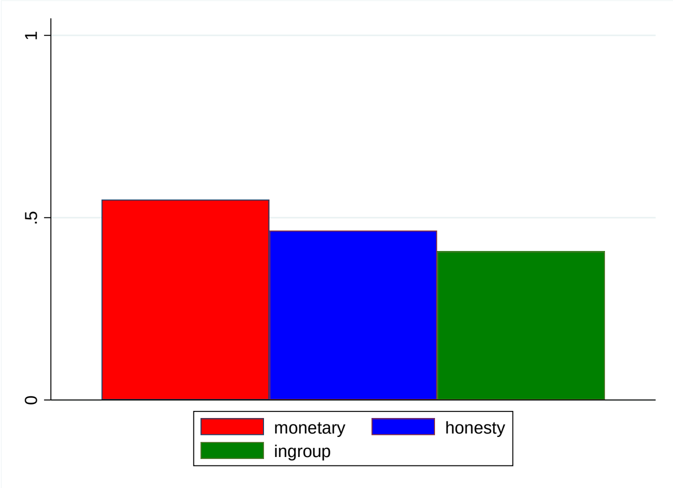
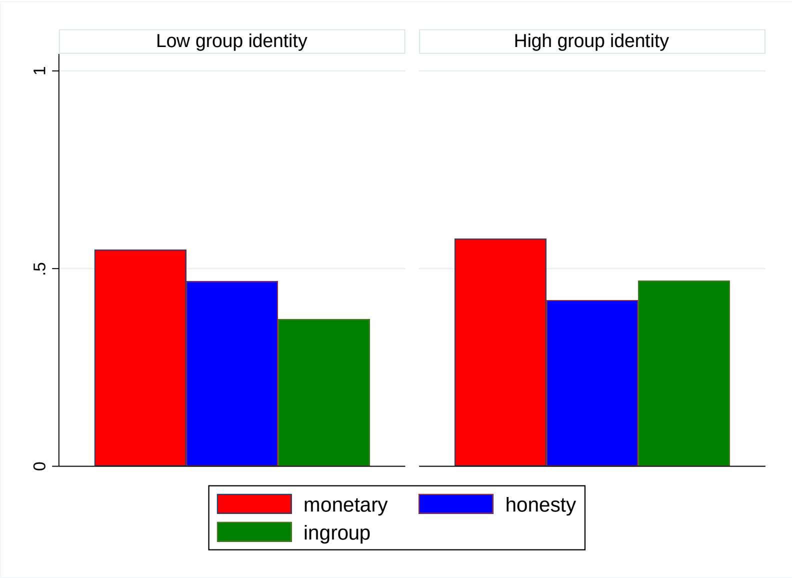


Figure 1 shows that the prevailing driver is monetary (in 54.97% of answers), but also honesty and pro-ingroup concerns play a relevant role (46.47% and 40.74%, respectively).[12] In addition, the data show that the presence of answers containing pro-ingroup concerns is positively correlated with monetary consideration (Spearman correlation test, with coefficient = .171, $p < 0.001$), but negatively correlated with honesty (Spearman correlation test, with coefficient = -.099, $p < 0.001$).

In fact, subjects reporting monetary-based motivations often refer to own but also ingroup's gains (*"I picked the option that would help my group more"*), while they justify their lies and their lower concern for honesty in the name of the group's sake (*"I want to maximize my group's reward so I deceived the yellow group member"*).

[12] Three nice example of answers that can be classified in the three categories are, respectively: "*I do these studies to feed my cats and myself and need to maximize the time spent on the platform as much as I can*", "*I chose the honest answer. Political party or not, I'm not here to lie to folks*", and "*I opted for the larger portion for my team*".

Figure 2. Group identity and percentage of different motivations for choice made.



An interesting result emerges if we compare subjects with high ingroup identity (those who report a level of identification that is equal or higher than the median, i.e., at least 6 in a 1-10 range) and low group identity (symmetrically, those who report a value lower than 6), see Figure 2. When motivating their choice, individuals reporting high ingroup identity are significantly more driven by pro-ingroup considerations (46.98% versus 37.20%: Z = -2.581, $p = 0.001$, two-sample Wilcoxon rank-sum test) and less by honesty, although this latter difference is not significant (42.04% versus 46.80%: Z = 1.255, $p = 0.210$, two-sample Wilcoxon rank-sum test).

Figure 3. Silence and percentage of different motivations for choice made.

Silence has a rather different effect, see Figure 3: introducing the silence option significantly increases the concern for honesty (49.43% versus 43.53%: Z = -2.353, p = 0.019, two-sample Wilcoxon rank-sum test), while reducing pro-ingroup considerations (37.55% versus 43.91%: Z = 2.581, p = 0.001, two-sample Wilcoxon rank-sum test). Monetary concerns do not differ by much. The role of silence as a "compromise option" clearly emerges from subjects' answers. Here we report two examples of answers that emphasize this perspective: "*I would not want to give them more money than me but also do not want to mislead them, so I would rather say nothing*", and "*Sending no message introduces no bias and also does not require me to be honest or lose out*".

3.5. Receiver behavior

This section briefly analyzes the behavior of the receiver subjects in the Yellow group, i.e. those who (may) receive the deceptive or truthful message from the Green senders and have to choose between two options (A and B) without knowing the payoffs associated. Table 7 below summarizes the percentage of Yellow receivers who selected Option B (the most favorable to the Green group) versus Option A (the most unfavorable to the Green group), distinguishing between those who received the truthful message (Message 1), those who received the untruthful message (Message 2), those who did not received any message, but were aware about the chance to receive a message from the sender (Silence), and those who had no information at all (NoInfo).

Table 7. Receivers' choice (percentage of Option B)

Info received	Average	Std. Err	Obs.
Truth	0.408	0.017	843
Lie	0.570	0.016	938
Silence	0.255	0.021	427
NoInfo	0.330	0.471	203

The table reports the percentage of Yellow receiver subjects who chooses Option B as depending on the information they receive: Message 1, that says that Option A is the more convenient for the Yellow group (truth); Message 2, that says that Option B is the more convenient for the Yellow group (lie), no message (Silence), no information (NoInfo).

The table shows that Green's message is effective in influencing Yellow's choice: Option B is selected significantly more frequently when the Green sender claims Option B is more profitable for the receiver (i.e. she lies) than when the Green sender claims Option A is more profitable or equally profitable (in case of white lies) for the receiver (i.e. she tells the truth) (57.0% versus 40.8%: Z = -10.808, $p < 0.001$, two-sample Wilcoxon rank-sum test) or when she sends no message (25.5%: Z = -6.838, $p < 0.001$, two-sample Wilcoxon rank-sum test). Interestingly, when receiving no messages subjects tend to choose Option A more frequently than Option B, and this is observed also in the NoInfo treatment, i.e. when subjects do not receive any form of information (not even that the sender could send them a message and decided not to do)[13]: only 25.5% of subjects choose Option B when receiving no message, and a slightly higher proportion is observed when no information is given: 33.0%: Z = -1.953, $p = 0.051$, two-sample Wilcoxon rank-sum test). The form of group assignment, i.e. at random or based on political orientation, does affect the frequency of choices in favor of Option B: partisan groups tend to trust the opponents' message significantly less frequently (the percentage of choices in favor of Option B are 54.9% versus 63.5%, Z = 2.270, $p = 0.023$, two-sample Wilcoxon rank-sum test). Furthermore, there is some difference between Democrats and Republicans in following the untruthful message when received (50.0% versus 59.9%: Z = 2.657, $p = 0.007$, two-sample Wilcoxon rank-sum test). Additionally, comparing subjects who strongly identify with their ingroup and subjects who do not, we find that strong identification with own Yellow group (defined as the case where the identification level is equal or higher than 6) is associated with a significantly higher tendency not to follow the message (Z = 3.141, $p = 0.002$, two-sample Wilcoxon rank-sum test). As with senders of blue lies, we observe receivers having a significantly higher ingroup identity when groups are formed depending on political orientation: 6.743 vs. 4.404, Z = 24.493, $p < 0.001$, two-sample Wilcoxon rank-sum test. Thus, partisanship does not determine *per se* the likelihood of untrusting the opponent, but raises the strength of ingroup identity, which then decreases trust towards the opponent group. Nonetheless, receivers with a high group identity still fall victim of senders' lies in 53.2% of cases (see Table A6 in the Appendix).

4. Structural model and estimation

To further understand the reduced-form results, to quantify the relevance of concurrent factors determining the choice of blue-lying, and to conduct a welfare analysis, we present a two-stage game between a potential blue-liar (sender) and her opponent (receiver). Since we are interested in studying the ingroup-outgroup dimension of a lie that goes in favor of the sender's ingroup and against the receiver's ingroup, the game we study involves four agents: the sender and her group peer (who, in the experiment, are called the "Greens"), and the receiver and her group peer (the "Yellows"). For sake of simplicity, we consider a situation where only the (Green) sender and the (Yellow) receiver

[13] The order in which we presented the two options A and B was randomized across respondents, thus the preference in favor of Option A cannot be motivated by the fact of being the "first" option. We can speculate that the preference for Option A could be explained by the positive connotation associated to the letter A with respect to B: for instance, A is associated to a higher rating or grade; having a "plan B" typically refers to a second best.

make decisions, while their peers are passive bystanders whose payoffs are affected by the sender's and the receiver's decisions. Although we believe the role of peers to be important in understanding how blue lies originate and spread, and thus worth of investigation, we leave such analysis to future work and concentrate on the sender and receiver.

The sender (Green). In the first stage, the sender decides whether to lie or not to the receiver and, in doing so, weighs three main concerns: monetary, pro-group and honesty. We denote with y the blue-lying choice. First, the blue lie is beneficial for the sender from a monetary point of view: the relevance of this monetary gain is captured by $\overline{\pi}^{Y} = \pi_B^G - \pi_A^G > 0$, where π_B^G is the Green sender's payoff if the Yellow receiver choses Option B (the more profitable for the Green group) and π_A^G is the Green sender's payoff if the Yellow receiver choses Option A (the less profitable for the Green group). Analogously, the Yellow receiver's payoffs will be denoted π_A^Y and π_B^Y, and the difference as $\overline{\pi}^{Y} = \pi_A^Y - \pi_B^Y$. However, gains are probabilistic, since they depend on whether the receiver believes the content of the message is true, and follows it. The probability the receiver makes the choice more favorable to the sender (allowing her to earn $\pi_B^G > \pi_A^G$) is denoted by h and depends on how the receiver reacts to the message she gets from the sender, as illustrated below.

Second, we assume the sender experiences an individual quadratic loss of lying since, as in other forms of lying, blue-lying brings intrinsic costs (Gneezy et al., 2018; Abeler et al., 2019) such as the moral cost of departing from honest behavior, whose size is captured by parameter c,and the cost of feeling guilty over the consequences of the lie g, since the lie can determine a lower payoff for the receiver and her ingroup mate. While blue lies entail both the moral cost of lying c and the cost of consequences g, white lies are characterized by c only.

Third, the sender may experience a pro-ingroup benefit, β, since she cares about the gains that (also) the ingroup's members can obtain from it.

Therefore, the sender choses the optimal lying content y in order to maximize the following utility:

$$U(y) = h(y)\pi_B^G + [1 - h(y)]\pi_A^G - \frac{(c+g)y^2}{2} + \beta y \qquad (1)$$

The pro-ingroup benefit β can be further specified by assuming it is dependent on two components, namely, the level of ingroup identity β_0, and the sensitivity β_1 to the ingroup's (versus the outgroup's) social norm n about blue-lying according to the following relation:

$$\beta = \beta_0 + \beta_1 n \qquad (2)$$

where $n = 1$ if the norm is blue-lying, and $n = -1$ if the norm is telling the truth. We denote with β_{1I} the subject's sensitivity to the ingroup (I) norm, while β_{1O} is the subject's sensitivity to the outgroup (O) norm. The ingroup norm is expected to engender conformity to ingroup behavior (if $n = -1$, then $\beta_{1I} > 0$; if $n = 1$, then $\beta_{1I} < 0$), while the outgroup norm is expected to be ignored (if $n = -1$ or if $n = 1$, then $\beta_{1O} = 0$).

The receiver (Yellow). The receiver enters the game being informed that she has to decide blindly between two options that give an unknown payoff to herself and her ingroup peers, and a different, unknown payoff (that can be higher or lower) to an outgroup also made of two people. The only information the receiver can get is a message from one of the two members of the outgroup (the sender), that may affect the probability of selecting the option that gives the outgroup the higher payoff ${\pi_A}^Y$, and the lower payoff ${\pi_B}^Y$ to herself and her ingroup peer. The sender, however, can decide not to send any message (as happens in the "silence" condition). If the receiver does not get any message, the probability to select the most unfavorable option for her and her ingroup is equal to a baseline probability $h_0 \in (0,1)$. If she receives the message, she can react by correcting this probability such as it rises (if the message is believed) or decreases (if it is not), thus adjusting the probability to $h \in [0,1]$ at a cost $k(h)$, with $k\,(h_0) = 0$, $k'(h_0) = 0$, and $k''(\cdot) > 0$. In fact, we assume the marginal cost of small adjustments as small, but larger adjustments have an increasing cost. The probability is adjusted of a factor ah, with $a = 0$ when the receiver does not get any message or is not influenced by its content, $a > 0$ if the receiver is (at least partially) convinced by the message, and $a < 0$ if she thinks that the message is deceitful and reacts selecting the opposite option. The receiver thus maximizes the following utility function:

$$U(h) = (h_0 + ah)\pi_B^Y + [1 - (h_0 + ah)]\,\pi_A^Y - k(h) \qquad (3)$$

Solving the model: the receiver (Yellow). We solve the model working backward, considering the receiver first. Identification of the structural parameters is achieved by assuming $k(h)$ to be quadratic and exploiting the variation from the experimental design between the receiver's choice when the untruthful message is received in the four experimental conditions that characterize the receiver (random versus partisan group formation, combined with the absence or presence of the silence option), and her respective choice in case of silence. We can define the following moment conditions:

$$E(h_{R-NS}) = a_{R-NS}\overline{\pi}^{Y}$$

$$E(h_{PO-NS}) = a_{PO-NS}\overline{\pi}^{Y}$$

$$E(h_{R-S}) = a_{R-S}\,\overline{\pi}^{Y}$$

$$E(h_{PO-S}) = a_{PO-S}\overline{\pi}^{Y}$$

where R refers to the Random condition, PO to the Political Orientation condition, NS to the NoSilence condition, and S to the Silence condition.

The system is overidentified and estimation of the model parameters is achieved via GMM. This implies minimizing the weighted distance between the empirical values of the adjustment choice in the NoSilence condition and that predicted by the model. The sources of identification for the probability adjustment when the blue lie is received (a_{R-NS}) and groups are formed randomly is E(h_{R-NS}), i.e., the adjustment in the probability of choosing Option B with respect to the probability of choosing Option B in the Silence condition when no message is received; the source of identification for the probability adjustment when the blue lie is received (a_{PO-NS}) and groups are based on political orientation is E(h_{PO-NS}), i.e., the adjustment in the probability of choosing Option B with respect to the probability of choosing Option B in the Silence condition when no message is received.

Table 8: Yellow receivers. Estimated preference parameters

VARIABLES	(1)	(2)
a_{R-NS}	0.190***	0.184***
	(0.016)	(0.025)
a_{PO-NS}	0.147***	0. 130***
	(0.009)	(0.011)
a_{R-S}	0.021	0.015
	(0.020)	(0.035)
a_{PO-S}	-0.009	-0.002
	(0.012)	(0.015)
Observations	1365	841
GMM Crit.	4.45e-33	9.12e-33

The Table reports the set of preference parameters (where a is the adjustment of probability that follows a possible message received) estimated using GMM. R refers to the Random condition, PO to the Political Orientation condition, NS to the NoSilence condition, and S to the Silence condition. Column (1) and (2) report the estimation results for the full model using the identity matrix as weighting matrix. Column (2) reports the estimation results for the restricted model (considering only subjects whose identification with the ingroup is equal or higher than 6).
*** significant at 1%; ** significant at 5%; * significant at 10%.

Our baseline estimation strategy uses an identity matrix as a weighting matrix[14], which overcomes the issue arising from the poor small sample properties of the optimal weighting matrix (see Altonji and Segal, 1996). Table 8 reports the estimated preference parameters for the model: the two simulated moments fall within the 95% confidence interval of the empirical moments. As a "sanity check", we also show that adjustment, as expected, is not significantly different from zero in the Silence condition (in principle, the receiver can derive some clue from receiving no message, but it appears that no relevant adjustment occurs if no message is received). Column 2 presents the estimation results for the restricted model in which we focus on subjects who have a level of group identity that is higher than the median value (which, as above, is 6, in a 1-10 range). The aim is focusing on how our observed mechanisms work in a context where people highly identify with their ingroup, as happens in "echo chambers" environments and alike. Consistently with our expectations, the probability adjustment that follows the message is smaller with partisan groups and for subjects highly identify with their ingroup, since both cases entail lower trust towards the outgroup.

Solving the model: the sender (Green). We now consider the sender's choice, whose optimal blue-lying choice from Equation (1) assumes the following form:

$$y^* = \frac{h'(y)\overline{\pi}^G + \beta_0 + \beta_1 n}{c + g}$$

The rate of blue lies increases in the monetary gains deriving by lying π_B, weighted for $h'(y)$, the strength of group identity β_0, and the sensitivity to the ingroup (vs. outgroup) social norm β_1.

The parameter $\beta_1$is further specified as β_{1I} when we consider the ingroup's norm, and as β_{1O} when we consider the outgroup's norm. Identification of the parameters $c^{NS}, g^{NS}, {\beta_0}^{NS}$, ${\beta_{1I}}^{NS}, {\beta_{1O}}^{NS},\ c^S, g^S, {\beta_0}^S, {\beta_{1I}}^S, {\beta_{1O}}^S$ is achieved by exploiting the features and the variation from the experimental design. We can define the following moment conditions:

$$\mathrm{E}(\mathrm{y}^W_{R-NON-NS}) = \frac{h'^{NS}\overline{\pi}^G}{c^{NS}}$$

$$\mathrm{E}(\mathrm{y}^W_{PO-NON-NS}) = \frac{h'^{NS}\overline{\pi}^G}{c^{NS}_{PO}}$$

[14] Unreported estimations show identical results with a diagonal matrix as weighting matrix.

$$E(y^B_{R-NON-NS}) = \frac{h'^{NS}\overline{\pi}}{c^{NS} + g^{NS}}$$

$$E(y^B_{PO-NON-NS}) = \frac{h'^{NS}\overline{\pi}^G + \beta_0{}^{NS}}{c^{NS}_{PO} + g^{NS}}$$

$$E(y^B_{PO-IN-NS}) = \frac{h'^{NS}\overline{\pi}^G + \beta_0{}^{NS} - \beta_{1I}{}^{NS} n}{c^{NS}_{PO} + g^{NS}}$$

$$E(y^B_{PO-OUT-NS}) = \frac{\overline{\pi}^G + \beta_0{}^{NS} - \beta_{1O}{}^{NS} n}{c^{NS}_{PO} + g^{NS}}$$

$$E(y^W_{R-NON-S}) = \frac{h'^{S}\overline{\pi}^G}{c^{S}}$$

$$E(y^W_{PO-NON-S}) = \frac{h'^{S}\overline{\pi}^G}{c^{NS}_{PO}}$$

$$E(y^B_{R-NON-S}) = \frac{h'^{S}\overline{\pi}^G}{c^{S} + g^{S}}$$

$$E(y^B_{PO-NON-S}) = \frac{h'^{S}\overline{\pi}^G + \beta_0{}^{S}}{c^{NS}_{PO} + g^{S}}$$

$$E(y^B_{PO-IN-S}) = \frac{h'^{S}\overline{\pi}^G + \beta_0{}^{S} - \beta_{1I}{}^{S} n}{c^{NS}_{PO} + g^{S}}$$

$$E(y^B_{PO-OUT-S}) = \frac{h'^{S}\overline{\pi}^G + \beta_0{}^{S} - \beta_{1O}{}^{S} n}{c^{NS}_{PO} + g^{S}}$$

where R refers to the Random condition, PO to the Political Orientation condition, NON to the NoNorm treatment, IN to the Ingroup norm treatment, OUT to the Outgroup norm treatment, NS to the NoSilence condition, and S to the Silence condition.

As above, the system is overidentified and estimation of the model parameters is achieved via GMM. This implies minimizing the weighted distance between the empirical value of the lying choices in each condition and those predicted by the model.

The source of identification for the pure moral cost of lying when silence is not available (c^{NS}) is $E(y^W_{R-NON-NS})$, i.e. the *white*-lying rate when groups are formed randomly, while the cost of lying in the same situation when groups are formed based on partisanship (c^{NS}_{PO}) is $E(y^W_{PO-NON-NS})$. The cost of feeling guilty over the consequences paid by the outgroup g^{NS} is identified by comparing white and blue lying rates when groups are formed randomly and no norms are communicated. The perceived degree of ingroup identification ($\beta_0{}^{NS}$) is identified by comparing the random and political orientation-based group formation. The identification of the weight attached to the ingroup norm ${\beta_{1I}}^{NS}$ and the outgroup norm ${\beta_{1O}}^{NS}$ relies on the difference in the lying rates when subjects receive the ingroup and the outgroup norm, respectively, and the lying rates in the absence of any norm. Finally, the identification of those parameters when silence is possible ($c^S, c_{PO}^{NS}, g^S, {\beta_0}^S, {\beta_{1I}}^S, {\beta_{1O}}^S$) relies on the comparison of treatments with and without silence.

Table 9: Senders. Estimated preference parameters

VARIABLES	(1)	(2)
c^{NS}	0.895***	0.802***
	(0.075)	(0.083)
c_{PO}^{NS}	0.928***	0.881***
	(0.055)	(0.056)
g^{NS}	0.372**	0.248
	(0.167)	(0.167)
${\beta_0}^{NS}$	-0.069	-0.022
	(0.067)	(0.087)
${\beta_{1I}}^{NS}$	-0.231***	-0.258***
	(0.047)	(0.054)
${\beta_{1O}}^{NS}$	0.059	0.011
	(0.050)	(0.058)
c^S	0.507***	0.459***
	(0.050)	(0.057)
c_{PO}^S	0.561***	0.555***
	(0.044)	(0.047)
g^S	0.363***	0.216
	(0.143)	(0.145)
${\beta_0}^S$	-0.034	0.041
	(0.055)	(0.074)
${\beta_{1I}}^S$	-0.088**	-0.061
	(0.044)	(0.049)
${\beta_{1O}}^S$	0.077	0.119***
	(0.047)	(0.054)
Observations	1,975	1,423
GMM Crit.	4.64e-13	5.59e-15

The Table reports the set of preference parameters estimated using GMM. NS refers to the NoSilence condition, and S to the Silence condition. Column (1) and (2) report the estimation results for the full model using the identity matrix as weighting matrix. Column (2) reports the estimation results for the restricted model (considering only subjects whose identification with the ingroup is equal or higher than 6).
*** significant at 1%; ** significant at 5%; * significant at 10%.

As for the receiver, our baseline estimation strategy in Table 9 uses an identity matrix as a weighting matrix[15]. Table 9 shows that all simulated moments fall within the 95% confidence interval of the empirical moments, with the exception of β_0 and β_{1O}.

As for Table 8, column 2 presents the estimation results for the restricted model in which we focus on subjects who have a level of group identity that is higher than the median value (which is 6, in a 1-10 range).

The estimation confirms that blue-lies decrease in the perceived moral cost from lying c, and shows that the possibility to remain silent reduces such loss considerably: in fact, $c^S < c^{NS}$ and $\frac{c^{NS}}{c^S} = 1.77$, meaning that allowing for silence – i.e. permitting not to intervene in favor of either the ingroup or the outgroup – reduces the relative moral cost of blue-lying of about 44% with respect to the No Silence condition. Such sizable blue-lying costs suggests that the welfare implications of allowing for silence can be large: not only silence increases the receiver's welfare, but also reduces the costs of lying bared by the sender when forced to decide between favoring the ingroup or damaging it (see Table 10 below for a more extensive discussion on welfare implications of blue lies).

Concerning the role of group identity, identifying with a group *per se* does not affect the tendency to lie in favor of the group (β_0 is not significantly different from 0), but through the sensitivity to ingroup norm β_{1I}. As expected, such sensitivity is positive (subjects' tendency to lie reduces when the norm is not lying, i.e. when $n = -1$) and stronger for subjects who highly identify with the group. In the No Silence condition (see Table 4 above), blue lies reduce from 43.2% to 28.8%, i.e., a drop of 33.3%. In the silence condition, this reduction is smaller (from 26.7% to 21.7%) and equal to 18.7% of the initial proportion. Silence reduces subjects' sensitivity to norms of more than 62%: in fact, $\frac{\beta_{1I}{}^{NS}}{\beta_{1I}{}^{S}} = 2.62$. Furthermore, it is worth noting that senders who strongly identify with their ingroup do not feel guilty for the monetary cost they impose to the Yellows when blue-lying (see in Table 9 that the parameter g is no more significant, no matter the possibility of remaining silent).

We now turn to welfare analysis of the effects of blue lies, with a specific attention to whether and how blue lies are detrimental for the society, and how the option of remaining silent can increase social welfare.

To assess the welfare effects of blue-lying, we structurally estimate the two-stage model parameters and use a minimum-distance estimator on the combined data from the receivers and the senders. The moments are the following: the receivers' probability adjustment when receiving the

[15] As for Table 8, unreported estimations show identical results with a diagonal matrix as weighting matrix.

message, the sender's costs of lying, and sender's sensitivities to the ingroup's and outgroup's social norms. Key parameters are the curvature of the cost of lying and cost of adjustment functions, that we assume to be both quadratic.

To have a comprehensive assessment of how blue lies affect social welfare, we consider as benchmark the situation where the receiver makes her choice with no information from the sender (NoInfo treatment). As shown in Table 7, the 33.0% of the receivers' choices go in favor of option A, generating an average payoff of $1.33 for the Green sender, and consequently an average payoff of $1.67 for the Yellow receiver. Since there is no exchange of information, and not even the information that derives from the lack of a message, there is no cost on both the opponents' side, and the social welfare of $3 for each pair.

As already shown in Table 4, silence reduces blue lies significantly: its consequence is a large reduction of receiver's choices in favor of the less favorable option to her (Option B). The obvious effect is a transfer of welfare from senders to receivers, who earn the highest payoff ($2) more frequently, at the expenses of the senders (who turn to the lowest payoff ($1). However, this is not the whole story. As shown in Table 9, the option of remaining silent partially sets off the senders' costs of lying, since it gives senders' the additional possibility of not acting either in favor of the ingroup or at its expenses, thus increasing the senders' welfare. Furthermore, as displayed in Table 8, receivers tend to adjust their expected payoffs once they get a message, although this is inefficient in case of an untruthful message whose information is detrimental to the receiver's choice. On the contrary, silence has been shown to entail no significant adjustment, and consequently no significant adjustment cost, thus increasing the receivers' welfare.

Panel A in Table 10 compares the utility (expressed in dollars) that derives to senders when choosing to blue-lie, telling the truth or remaining silent. Having the chance to communicate, giving or receiving some information, entails a utility loss with respect to the benchmark no matter the sender's choice. Comparing the three choices, as expected blue-lying generally implies a higher utility for senders with respect to truth-telling: the higher payoff more than compensates the costs of lying. Remaining silent is on average more beneficial than blue-lying for senders, since it rules out lying costs without monetarily damaging the individual (ad its ingroup) in favor of the outgroup.

Panel B presents the utility that accrues to receivers when being victim of lies instead of getting a truthful message or no message at all. Again, there is a utility loss with respect to the benchmark. As expected, no matter the treatment and the condition, blue-lying always imply a lower utility for receivers with respect to truth. This (negative) difference more than compensates the increase in utility that senders may gain, showing that blue lies are detrimental for the society.

In addition, receivers' utility is also larger when silence is available, being this due to the adjustment costs they would bear when receiving a message.

Table 10. Welfare analysis of blue lies: utility comparison

Panel A: Senders' utility

Benchmark (U^{NI}): 1.330

	Blue lies	Truth	Silence
$U^{B-R-NON}$	0.973	0.821	1.101
	(0.003)	(0.178)	(0.021)
$U^{B-PO-NON}$	0.970	0.797	1.104
	(0.002)	(0.100)	(0.015)
$U^{B-PO-IN}$	0.924	0.893	1.067
	(0.009)	(0.079)	(0.013)
$U^{B-PO-OUT}$	0.966	0.773	1.071
	(0.002)	(0.103)	(0.008)
Observations	681	681	705
GMM Crit.	2.03e-31	2.13e-32	3.19e31

Panel B: Receivers' utility

Benchmark (U^{NI}): 1.670

	Blue lies	Truth	Silence
$U^{B-R-NON}$	0.895	1.214	1.405
	(0.009)	(0.023)	(0.080)
$U^{B-PO-NON}$	0.910	1.300	1.527
	(0.008)	(0.015)	(0.049)
Observations	427	843	427
GMM Crit.	1.32e-31	1.88e-32	6.72e-33

The Table reports the utility in dollars derived from blue-lying, truth-telling, or remaining silent estimated using GMM for senders (Panel A) and receivers (Panel B). B refers to blue lies. R refers to the Random condition, PO to the Political Orientation condition, NON to the NoNorm treatment, IN to the Ingroup norm treatment, OUT to the Outgroup norm treatment, NI to the NoInfo condition. We restrict the analysis to the Silence condition, in which senders have the possibility to choose among the three alternative options.

Overall, both groups would have been better without exchanging any information. But since the Greens are in the position of conveying a message, they tend to blue-lie, especially in highly partisan environments, nonetheless paying the costs of these lies. Receivers are subject to these lies, that not only reduce their monetary payoff, but also generate a cost of updating their beliefs that is fully inefficient. Being able to increase both senders' and receivers' utility, silence has therefore a social welfare enhancing effect since it discourages blue lies in favor of truth and does not require any adjustment costs on the receivers' side.

As we will discuss more extensively below, this result sheds light on possible directions of intervention to limit blue lies in various contexts, and helps also understanding some of the reasons why omertous behavior has imposed for centuries as the prevailing norm in societies characterized by institutional dualism (mafia vs. state).

Interestingly, the welfare analysis for white lies shows that white-lying, differently from blue-lying, is almost as beneficial as remaining silent, see Table A7.

5. Discussion and conclusion

Blue lies are a pervasive phenomenon. Our findings show that the rate of blue lies is substantial, ranging between 38% and 57%. Respondents lie for reasons that apply to both themselves *and their ingroup*: lying pays more money than telling the truth, although they acknowledge that lying has a moral cost. Partisan behavior comes at the expenses of honesty; the importance subjects ascribe to honesty shrinks further when ingroup identity is strong – which happens when group formation is based on a highly polarizing issue such as political orientation.

But why should we worry about blue lies? As shown by our welfare analysis, blue lies reduce social welfare significantly. Since the applications of the blue lies game are many, spanning from fake news, echo chambers, and hate speech phenomena to the risk of manipulating electoral outcomes and putting democratic institutions in danger, the extent of the social loss involved is remarkable. In recent years, the growth of social media interactions and the increasing consumption of online news, often transmitted through social networks, has favored the diffusion of content among like-minded individuals, forming echo chambers where people are insulated from the contrary perspective and third-party fact checking or editorial judgement (Allcott and Gentzkow, 2019) is not available. Accuracy is costly; furthermore, news consumers enjoy and feel reassured by partisan views. Verifying (and then telling) the truth has the double cost of requiring effort and challenging partisanship (Gneezy et al., 2012). Spreading fake news means speaking to one's own ingroup, affirming common values and, potentially, promulgating conspiracy theories, hate, xenophobia, and prejudice (Ameur and Aliane, 2021). These phenomena are getting quite worrying now that social media have acquired increasing impact on people's daily life, affecting their wellbeing and influencing attitudes, voting participation and outcomes (Long et al., 2019)

The introduction of the blue lie game helps us to understand why blue lies are told, which mechanisms they rely on, and how to challenge them. Our result show that highly polarizing issues, such as political leaning, engender partisanship and, although a stronger sense of group identity does not increase blue lies *per se*, it affects how people react to lying norms, potentially leading to a dramatic increase of blue lies when blue lies are a prevalent choice within the ingroup, as shown in the analysis based on the structural estimation of our model parameters. Counteraction is needed because blue lies reduce social welfare significantly.

Our results suggest two possible lines of intervention. First, manifesting a social norm of truth-telling has a wide and significant impact on the reduction of blue lies, especially for subjects who highly identify with the group. Our results show that high group identity comes with higher willingness to differentiate from the outgroup. This translates into promoting fact checking and debunking *per se*, but particularly within a circle people feel they belong. The "good example" must be given by people with whom we strongly identify. With weaker identification, the effect is smaller, and when expressed by people perceived as distant (or opponents) might even backfire.

Second, changing the architecture of the choice by adding the option of remaining silent has a considerable effect in reducing blue lies. Making salient the possibility to be silent (for instance by asking whether the person really wishes to comment and share content) not only induces people to double think before speaking, but it also makes clear that expressing extreme and polarized opinions is not the only way to handle a discussion. This result emphasizes the damages perpetrated by current social media algorithms, which reward participation and promote any expression of opinion, only intervening ex post by limiting hate speech and offensive comments when the storm has already become viral.

If remaining silent can be a good alternative to tell a lie, it still implies not telling the truth. In Mafia-like organizations, omertous behavior is at the foundation of Mafia's code of honor and testifies the bond that holds the members of the Mafia brotherhood together. The Mafioso who does not abide by the code of silence is expelled from the association and must be killed. In this respect, "larger doses of *omertà* ensure lesser ones of violence" (Gambetta, 2000, p.170), and might be welfare-enhancing as shown in our model.

Given the novelty of this paper, we hope and believe it may open several avenues of future research, from testing the effect of building groups based on polarizing issues other than politics, to manipulating how information is shared within the group before and after the blue-lying decision, to allowing for the possibility to block lies within the group before they emerge, and to exploring the benefits and costs of silence, among others. Given the relevance of the blue lies phenomenon, we expect many further developments will follow.

References

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. Econometrica, 87(4), 1115-1153.

Aksoy, B., & Palma, M. A. (2019). The effects of scarcity on cheating and in-group favoritism. Journal of Economic Behavior & Organization, 165, 100-117.

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of economic perspectives, 31(2), 211-236.

Allcott, H., & Kessler, J. B. (2019). The welfare effects of nudges: A case study of energy use social comparisons. American Economic Journal: Applied Economics, 11(1), 236-276.

Altonji, J. G., & Segal, L. M. (1996). Small-sample bias in GMM estimation of covariance structures. Journal of Business & Economic Statistics, 14(3), 353-366.

Ameur, M. S. H., & Aliane, H. (2021). AraCOVID19-MFH: Arabic COVID-19 multi-label fake news & hate speech detection dataset. Procedia Computer Science, 189, 232-241.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., ... & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences, 115(37), 9216-9221.

Barnes, J. A. (1994). A pack of lies: Towards a sociology of lying. Cambridge University Press.

Berger, J., & Heath, C. (2008). Who drives divergence? Identity signaling, outgroup dissimilarity, and the abandonment of cultural tastes. Journal of Personality and Social Psychology, 95(3), 593–607. https://doi.org/10.1037/0022-3514.95.3.593

Benard, S. (2012). Cohesion from conflict: Does intergroup conflict motivate intragroup norm enforcement and support for centralized leadership?. Social Psychology Quarterly, 75(2), 107-130.

Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. Political Psychology, 41, 3-33.

Boldt, L., & Arora, N. (2017). Dyadic compromise effect. Marketing Science, 36(3), 436-452.

Bok, S. (1978). Lying: Moral choice in public and private life. Random House; New York.

Borsellino, C. (2013). Motivations, moral components, and detections of lying behavior to benefit self and others. International Journal of Psychology and Behavioral Sciences, 3(3), 70-76.

Bursztyn, L., Egorov, G., & Fiorin, S. (2020a). From extreme to mainstream: The erosion of social norms. American economic review, 110(11), 3522-3548.

Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2020b). Misperceived social norms: Women working outside the home in Saudi Arabia. American economic review, 110(10), 2997-3029.

Cadsby, C. B., Du, N., & Song, F. (2016). In-group favoritism and moral decision-making. Journal of Economic Behavior & Organization, 128, 59-71.

Carl, T., & Bussey, K. (2022). Cross-sectional and longitudinal relationships between children's moral standards and their antisocial lie telling. Journal of Applied Developmental Psychology, 80, 101411.

Charness, G., & Chen, Y. (2020). Social identity, group behavior, and teams. Annual Review of Economics, 12, 691-713.

Charness, G., Cooper, D. J., & Grossman, Z. (2020). Silence is golden: team problem solving and communication costs. Experimental Economics, 23, 668-693.

Charness, G. Rigotti, L., & Rustichini, A. (2007). Individual behavior and group membership. American Economic Review, 97, 1340-1352)

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. American Economic Review, 99(1), 431-457.

Chopra, F., Haaland, I., & Roth, C. (2023). The demand for news: Accuracy concerns versus belief confirmation motives. NHH Dept. of Economics Discussion Paper, (01).

de Arruda, H. F., Cardoso, F. M., de Arruda, G. F., Hernández, A. R., da Fontoura Costa, L., & Moreno, Y. (2022). Modelling how social network algorithms can influence opinion polarization. Information Sciences, 588, 265-278.

Dimant, E. (2024). Hate trumps love: The impact of political polarization on social preferences. Management Science, 70(1), 1-31.

Dimant, E., Gee, X; Krupka, E., & Robbett, A. (2023). Escalation and polarization spillovers. Mimeo.

Dimant, E., Gelfand, M. J., Hochleitner, A., & Sonderegger, S. (2023). Strategic behavior with tight, loose, and polarized norms. Loose, and Polarized Norms (April 2, 2023).

Ellemers, N., Van Der Toorn, J., Paunov, Y., & Van Leeuwen, T. (2019). The psychology of morality: A review and analysis of empirical studies published from 1940 through 2017. Personality and Social Psychology Review, 23(4), 332-366. https://doi.org/10.1177%2F1088868318811759

Erat, S., & Gneezy, U. (2012). White lies. Management Science, 58(4), 723-733.

Farhall, K., Carson, A., Wright, S., Gibbons, A., & Lukamto, W. (2019). Political elites' use of fake news discourse across communications platforms. International Journal of Communication, 13.

Fisher, R. J. (2000). Intergroup conflict. The handbook of conflict resolution: Theory and practice, 166-184.

Foertsch, M., & Roesel, F. (2023). Sports Clubs and Populism: Quasi-Experimental Evidence from German Cities. CESIFO Working Paper Series 10259.

Foucart, R., & Tan, J. H. (2023). A test of loyalty. Theory and Decision, 1-29.

Fu, G., Luo, Y. C., Heyman, G. D., Wang, B., Cameron, C. A., & Lee, K. (2016). Moral evaluations of lying for one's own group. Infant and Child Development, 25(5), 355-370.

Fu, G., Xu, F., Cameron, C. A., Heyman, G., & Lee, K. (2007). Cross-cultural differences in children's choices, categorizations, and evaluations of truths and lies. Developmental psychology, 43(2), 278.

Fu, G., Evans, A. D., Wang, L., & Lee, K. (2008). Lying in the name of the collective good: A developmental study. Developmental science, 11(4), 495-503.

Gambetta, D. (2000). Mafia: the price of distrust. Trust: Making and breaking cooperative relations, 10, 158-175.

Gentzkow, M., & Shapiro, J. M. (2006). Media bias and reputation. Journal of Political Economy, 114(2), 280-316.

Gneezy, A., Imas, A., Brown, A., Nelson, L. D., & Norton, M. I. (2012). Paying to be nice: Consistency and costly prosocial behavior. Management Science, 58(1), 179-187.

Gneezy, U. (2005). Deception: The role of consequences. American Economic Review, 95(1), 384-394.

Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. Journal of Public Economics, 148, 14-31.

Halpern, D., & Gibbs, J. (2013). Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. Computers in human behavior, 29(3), 1159-1168.

Hornstein, H. A. (1976). Cruelty and kindness: A new look at aggression and altruism. Englewood Cliffs, NJ: Prentice Hall, Inc.

Hornsey, M. J., van Leeuwen, E. A. C., & van Santen, W. (2003). Dragging down and dragging up: How relative group status affects responses to common fate. Group Dynamics: Theory, Research, and Practice, 7, 275-288.

Hornung, M. (2016). Classifying prosocial lies: An empirical approach. International Review of Pragmatics, 8(2), 219-246.

Ikeda, A., Okumura, Y., Kobayashi, T., & Itakura, S. (2018). Children passively allow other's rule violations in cooperative situations. Scientific Reports, 8(1), 6843.

Lane, T., Miller, L., & Rodriguez, I. (2024). The normative permissiveness of political partyism. European Economic Review, 162, 104661.

Lau, Y. L., Cameron, C. A., Chieh, K. M., O'Leary, J., Fu, G., & Lee, K. (2013). Cultural differences in moral justifications enhance understanding of Chinese and Canadian children's moral decisions. Journal of Cross-Cultural Psychology, 44(3), 461-477.

Leib, M., Köbis, N., Soraperra, I., Weisel, O., & Shalvi, S. (2021). Collaborative dishonesty: A meta-analytic review. Psychological Bulletin, 147(12), 1241.

Levine, E., Hart, J., Moore, K., Rubin, E., Yadav, K., & Halpern, S. (2018). The surprising costs of silence: Asymmetric preferences for prosocial lies of commission and omission. Journal of personality and social psychology, 114(1), 29.

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. Organizational Behavior and Human Decision Processes, 126, 88-106.

Long, N. V., Richardson, M., & Stähler, F. (2019). Media, fake news, and debunking. Economic Record, 95(310), 312-324.

Mannino, G., Giunta, S., Buccafusca, S., Cannizzaro, G., & Lo Verso, G. (2015). Communication strategies in Cosa Nostra: an empirical research. World Futures, 71(5-8), 153-172.

McNair, B. (2018). Fake news: Falsehood, fabrication and fantasy in journalism. Oxon, UK: Routledge.

Mullainathan, S., & Shleifer, A. (2005). The market for news. American economic review, 95(4), 1031-1053.

Nezlek, J. B., & Smith, C. V. (2005). Social identity in daily social interaction. Self and Identity, 4, 243e261. http://dx.doi.org/10.1080/13576500444000308.

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. Episteme, 17(2), 141-161.

Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. Journal of Behavioral and Experimental Finance, 17, 22-27.

Pariser, E. (2011). The Filter Bubble: What the Internet Is Hiding from You. Penguin UK.

Shayo, M. (2020). Social identity and economic policy. Annual Review of Economics, 12, 355-389.

Sheng, S., Parker, A. M., & Nakamoto, K. (2005). Understanding the mechanism and determinants of compromise effects. Psychology & Marketing, 22(7), 591-609.

Shogren, J. (1989). Fairness in bargaining requires a context. Economics Letters, 31, 319-323.

Siegel J, Dubrovsky V, Kiesler S, et al. (1986) Group processes in computer-mediated communication. Organizational Behavior and Human Decision Processes 37(2): 157–187.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. Journal of consumer research, 16(2), 158-174.

Sio, U. N., Kotovsky, K., & Cagan, J. (2018). Silence is golden: The effect of verbalization on group performance. Journal of Experimental Psychology: General, 147(6), 939.

Smith, J. A. (2017). How the Science of "Blue Lies" May Explain Trump's Support. Scientific American Mind, 28(4), 77-80.

Sweet, M. A., Heyman, G. D., Fu, G., & Lee, K. (2010). Are there limits to collectivism? Culture and children's reasoning about lying to conceal a group transgression. Infant and Child Development, 19(4), 422-442.

Sunstein, C. R. (2001). Echo Chambers: Bush v. Gore, Impeachment, and Beyond. Princeton University Press.

Sunstein, C.R. (2007). Republic.com 2.0. Princeton University Press.

Tajfel, H. (1981). Human groups and social categories: Studies in social psychology. Cambridge, England: Cambridge University Press.

Sznajd-Weron, K., & Sznajd, J. (2000). Opinion evolution in closed community. International Journal of Modern Physics C, 11(06), 1157-1165.

Tajfel, H., Billig, M., Bundy, R. P., & Flament, C. (1971). Social categorization andintergroup behavior. European Journal of Social Psychology, 1, 149e178.

Theodoridis, A. G. (2017). Me, myself, and (I),(D), or (R)? Partisanship and political cognition through the lens of implicit identity. The Journal of Politics, 79(4), 1253-1267.

Van Bavel, J. J., & Packer, D. J. (2021). The power of us: Harnessing our shared identities to improve performance, increase cooperation, and promote social harmony. Little, Brown Spark.

Van Bavel, J. J., Packer, D., Ray, J. L., Robertson, C., & Ungson, N. D. (2022). How social identity tunes moral cognition. https://doi.org/10.31234/osf.io/9efsb

Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., & Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? A critical review. Social Issues and Policy Review, 11(1), 274-302.

Wu, T. Y., & Atkin, D. J. (2018). To comment or not to comment: Examining the influences of anonymity and social support on one's willingness to express in online news discussions. New Media & Society, 20(12), 4512-4532.

Appendix

Table A1: Experimental sessions – Green (Senders)

Session	Type of lie	Order	Group_formation	Treatment	Obs
1	BLUE	POST	RANDOM	NO NORM	200
2	BLUE	POST	PO Dem	NO NORM	201
3	BLUE	POST	PO Rep	NO NORM	199
4	BLUE	PRE	PO Dem	NO NORM	202
5	BLUE	PRE	PO Rep	NO NORM	198
6	BLUE	PRE	RANDOM	NO NORM	203
7	BLUE	PRE	PO Dem	INGROUP NORM	198
8	BLUE	PRE	PO Rep	INGROUP NORM	204
9	BLUE	POST	PO Rep	INGROUP NORM	203
10	BLUE	PRE	PO Dem	OUTGROUP NORM	199
11	BLUE	POST	PO Dem	OUTGROUP NORM	199
12	BLUE	PRE	PO Rep	OUTGROUP NORM	202
13	BLUE	POST	PO Rep	OUTGROUP NORM	203
14	BLUE	POST	PO Dem	INGROUP NORM	199
15	BLUE	PRE	RANDOM	INGROUP NORM	204
16	BLUE	POST	RANDOM	INGROUP NORM	199
17	WHITE	PRE	RANDOM	NO NORM	204
18	WHITE	PRE	PO Dem	NO NORM	201
19	WHITE	PRE	PO Rep	NO NORM	203
20	WHITE	POST	RANDOM	NO NORM	199
21	WHITE	POST	PO Dem	NO NORM	198
22	WHITE	POST	PO Rep	NO NORM	200
23	NOINFO	n.a.	RANDOM	NO NORM	208
24	NOINFO	n.a.	RANDOM	NO NORM	209

Table A2: Experimental sessions – Yellow (Receivers)

Session	Type of lie	Group_formation	Obs
1	BLUE	RANDOM	199
2	BLUE	PO Rep	199
3	BLUE	PO Dem	199
4	BLUE	PO Rep	202
5	BLUE	PO Dem	208
6	BLUE	RANDOM	199
7	BLUE	PO Rep	205
8	BLUE	PO Dem	198
9	BLUE	PO Dem	200
10	BLUE	PO Rep	202
11	BLUE	PO Rep	203
12	BLUE	PO Dem	203
13	BLUE	PO Dem	201
14	BLUE	PO Rep	198
15	BLUE	RANDOM	201
16	BLUE	RANDOM	200
17	WHITE	RANDOM	202
18	WHITE	PO Rep	202
19	WHITE	PO Dem	205
20	WHITE	RANDOM	202
21	WHITE	PO Rep	201
22	WHITE	PO Dem	201
23	NO INFO	RANDOM	203
24	NO INFO	RANDOM	202

Table A3. Frequency of white lies and group identity

Condition	Order	Average	Std. Err	Obs.
Random	PRE	0.565	0.074	46
PO Dem	PRE	0.654	0.067	52
PO Rep	PRE	0.471	0.071	51
Random	POST	0.659	0.072	44
PO Dem	POST	0.583	0.072	48
PO Rep	POST	0.659	0.072	44

The table reports subjects' white lies choices. Random refers to groups' random assignment, PO to groups' assignment based on Political Orientation. Dem refers to subjects who pre-screened as Democrats, while Rep refers to subjects who pre-screened as Republicans. PRE refers to sessions where the question on group identity was asked before the white-lying choice, while POST refers to sessions where the question on group identity was asked after the white-lying choice.

Table A4. Frequency of white lies and silence

Condition/treatment	Silence availability	Average	Std. Err	Obs.
Random	No silence	0.611	0.052	90
PO	No silence	0.590	0.035	195
Random	Silence	0.495	0.049	105
PO	Silence	0.447	0.035	199

The table reports subjects' white lies choices. Random refers to groups' random assignment, PO to groups' assignment based on Political Orientation. "No silence" corresponds to the condition when subjects can decide only between white-lying or telling the truth, whereas "Silence" corresponds to the condition when subjects can white-lie, tell the truth or remain silent.

Table A5: Determinants of lies: silence and norms

VARIABLES	(1) identity	(2) id&silence	(3) silence&ingroup norm	(4) silence&outgroup norm	(5) all	(6) MHT
white	0.370***	0.369***	0.373***	0.369***	0.372***	0.379***
	[0.074]	[0.074]	[0.074]	[0.074]	[0.074]	[0.075]
pol_orient	-0.094	-0.093	-0.099	-0.093	-0.099	-0.280***
	[0.076]	[0.076]	[0.076]	[0.076]	[0.076]	[0.077]
norm_own	-0.375***	-0.375***	-0.508***	-0.374***	-0.508***	-0.545***
	[0.077]	[0.077]	[0.098]	[0.077]	[0.100]	[0.101]
norm_others	0.223***	0.223***	0.227***	0.274**	0.237**	0.191*
	[0.086]	[0.086]	[0.086]	[0.111]	[0.113]	[0.110]
silence	-0.332***	-0.379**	-0.403***	-0.313***	-0.463***	-0.559***
	[0.056]	[0.158]	[0.064]	[0.062]	[0.165]	[0.172]
ident_level	0.017	0.014	0.017	0.017	0.012	0.010
	[0.011]	[0.015]	[0.011]	[0.011]	[0.016]	[0.016]
pre_id	-0.116**	-0.116**	-0.114**	-0.115**	-0.114**	-0.124**
	[0.056]	[0.056]	[0.056]	[0.056]	[0.056]	[0.056]
dem	0.057	0.057	0.056	0.057	0.055	0.023
	[0.065]	[0.065]	[0.065]	[0.065]	[0.065]	[0.065]
id_silence		0.007			0.010	0.022
		[0.022]			[0.022]	[0.023]
sil_own_norm			0.283**		0.281**	0.330**
			[0.129]		[0.134]	[0.138]
sil_others_norm				-0.102	-0.020	0.139
				[0.142]	[0.148]	[0.146]
Constant	-0.144	-0.121	-0.104	-0.152	-0.074	0.092
	[0.100]	[0.123]	[0.102]	[0.101]	[0.126]	[0.133]
Observations	2,174	2,174	2,174	2,174	2,174	2,137

Probit (standard errors in parentheses). The dependent variable is a dummy variable assuming value equal to 1 when the subject lies (being it a blue or a white lie), and 0 when she tells the truth. white is a dummy variable assuming value equal to 1 when the lie is white, and 0 when the lie is blue. pol_orient is a dummy variable assuming value equal to 1 when the groups are assigned depending on a prescreening based on political orientation, and 0 when group assignment occurs at random. norm_own is a dummy variable assuming value equal to 1 when the subject receives information on the descriptive ingroup norm, and 0 when she receives no information. norm_others is a dummy variable assuming value equal to 1 when the subject receives information on the descriptive outgroup norm, and 0 when she receives no information. silence is a dummy variable assuming value equal to 1 when the subject has the option of being silent on top of lying and telling the truth, and 0 when she has no possibility of being silent. ident_level is a variable ranging between 1 and 10 that captures how much a subjects identifies with the group she is assigned to. pre-id is a dummy variable assuming value equal to 1 when the question about the level of group identification is asked before the lying decision, and 0 when it is asked after the lying decision. dem is a dummy variable that assumes value equal to 1 when the subject has prescreened herself as "Democrat", and 0 if not. id_silence is the interacted variable between ident_level and silence. sil_own_norm is the interacted variable between silence and norm_own. sil_others_norm is the interacted variable between silence and others_norm.
*** significant at 1%; ** significant at 5%; * significant at 10%.

Table A6. Receivers' choice (percentage of Option B), restricted to high identity receivers

Info received	Average	Std. Err	Obs.
Message 1 (Truth)	0.433	0.021	543
Message 2 (Lie)	0.532	0.020	596
No message (Silence)	0.257	0.028	245

The table reports the percentage of subjects who chooses Option B as depending on the information they receive: Message 1, that says that Option A is the more convenient for the Yellow group (truth); Message 2, that says that Option B is the more convenient for the Yellow group (lie), no message (Silence).

Table A7. Welfare analysis of white lies: utility comparison

Panel A: Senders' utility

Benchmark (U^{NI}): 1.495

	White lies	Truth	Silence
$U^{W-R-NON}$	1.029	0.892	1.040
	(0.002)	(0.022)	(0.009)
$U^{W-PO-NON}$	0.990	0.856	1.038
	(0.001)	(0.017)	(0.005)
Observations	285	163	334
GMM Crit.	1.75e-30	4.35e-32	3.45e31

Panel B: Receivers' utility

Benchmark (U^{NL}): 2.000

	White lies	Truth	Silence
$U^{W-R-NON}$	1.794	1.893	1.894
	(0.011)	(0.014)	(0.019)
$U^{W-PO-NON}$	1.820	1.881	1.914
	(0.009)	(0.010)	(0.011)
Observations	265	216	119
GMM Crit.	2.75e-31	2.38e-31	7.65e-33

The Table reports the utility in euro derived from white-lying, truth-telling, or remaining silent estimated using GMM for senders (Panel A) and receivers (Panel B). W refers to white lies. R refers to the Random condition, PO to the Political Orientation condition, NON to the NoNorm treatment, NI to the NoInfo condition. We restrict the analysis to the Silence condition, in which senders have the possibility to choose among the three alternative options.