Nonverbal content and trust: An experiment on digital communication

Zakaria Babutsidze, Nobuyuki Hanaki, Adam Zylbersztejn

Nonverbal content and trust: An experiment on digital communication *

Zakaria Babutsidze† Nobuyuki Hanaki‡ Adam Zylbersztejn§ ¶

April 12, 2021

Abstract

We experimentally study the effect of the mode of digital communication on the emergence of trust in a principal-agent relationship. We consider three modes of communication that differ in the capacity to transmit nonverbal content: plain text, audio, and video. Communication is pre-play, one-way, and unrestricted, but its verbal content is homogenized across treatments. Overall, both audio and video messages have a positive (and similar) effect on trust as compared to plain text; however, the magnitude of these effects depends on the verbal content of agent's message (promise to act trustworthily vs. no such promise). In all conditions, we observe a positive effect of the agent's promise on the principal's trust. We also report that trust in female principals is sensitive to the availability of nonverbal cues about their partners.

Keywords: Digital communication; Trust; Hidden action; Nonverbal content; Principal-agent relationship; Promises.

JEL Code: C72, C92, D83

*The authors acknowledge the support from the following programs operated by the French National Research Agency (Agence Nationale de Recherche): DigiCom as a part of UCA^{JEDI} (ANR-15-IDEX-01) and LABEX CORTEX (ANR-11-LABX-0042) as a part of Université de Lyon (ANR-11-IDEX-007), as well as the Joint Usage/Research Center at ISER, Osaka University, and Grant-in-aid for Scientific Research, Japan Society for the Promotion of Science (15H05728, 18K19954, 20H05631).

†Associate professor, SKEMA Business School, Université Côte d'Azur, GREDEG, France and OFCE, Sciences Po Paris, Paris, France

‡Professor, Osaka University, Institute of Social and Economic Research, Ibaraki, Osaka, Japan

§Associate professor, Univ Lyon, Université Lyon 2, GATE UMR 5824, F-69130 Ecully, France and research fellow at Vistula University Warsaw (AFiBV), Warsaw, Poland

¶Corresponding author: zylbersztejn@gate.cnrs.fr

1 Introduction

Trust is of utmost value for economic activities (Rousseau et al., 1998). It can be defined as "the belief that others act in the interest of some measure of fairness or social welfare rather than their own self-interest" (Bracht and Feltovich, 2009, p. 1036).[1] Various institutions, spanning from centralized justice and law enforcement systems to self-governance within a community, foster and secure trust in economic interactions by means of incentives and punishment (Farrell and Knight, 2003). But trust may also emerge through tacit social agreements and informal norms of behavior (Greif, 1993). Such informal trust often hinges on reputation which, in turn, requires repeated interactions and ways of providing and receiving feedback (Bolton et al., 2004; Bracht and Feltovich, 2009; Simpson et al., 2017). Importantly, however, even when reputation lacks relevance, is hard to build, or when repeated interactions are unlikely, people may still exhibit a specific form of trust towards others – known as "swift trust" (Meyerson et al., 1996). Swift trust arises during a short process of interaction between previously unacquainted parties and (temporarily) affects their behavior.[2]

Recent experimental evidence suggests that face-to-face communication can be highly conducive in establishing trust in social interaction by facilitating the detection of other's trustworthiness (He et al., 2016). Albeit most primal, this form of communication is multilayer and encompasses not only verbal components (e.g., verbal content and style of the message), but also on nonverbal components (e.g., a speaker's facial displays, voice, body movements).[3] However,

[1] As they note, trust and trustworthiness (i.e., the extent to which trust in a person is warranted) are commonly considered as primary components of social capital (see also Knack and Keefer, 1997).

[2] Swift trust differs from the general willingness to trust other people – an other-regarding preference that can be measured either in behavioral or attitudinal manner (or both). A commonplace methodology for measuring behavioral trust in laboratory settings is the incentivized, one-shot and anonymous trust game based on Berg et al. (1995). In this context, trust is defined as "a willingness to bet that another person will reciprocate a risky move (at a cost to themselves)" (Camerer, 2003, p. 85) and captured by the amount transferred from the trustor to the trustee; accordingly, trustworthiness is the amount subsequently re-transfered in the opposite direction. See Johnson and Mislin (2011) for a meta-analysis of trust game experiments. Attitudinal trust, in turn, is often measured in large-scale surveys using items such as "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" (World Values Survey; see, e.g., Knack and Keefer, 1997), "In general, one can trust other people" (German Socio-Economic Panel, also used herein), or "I assume that people have only the best intentions" (Global Preference Survey being part of the Gallup World Poll; see Falk et al., 2018). In addition, Falk et al. (2016) provide evidence that the two types of measures are correlated.

[3] As noted by Chovil and Fridlund (1991), "[f]acial displays are a means by which we communicate with others. Like words and utterances, they are more likely to be emitted when there is a potential recipient, when they are

the ongoing IT revolution that has been increasingly replacing direct face-to-face interactions with those mediated by new digital technologies. Many strangers are now transacting in open online marketplaces without ever meeting in person. People are donating to charity online (e.g. *Qgiv*) or contribute to online crowd-funding campaigns (e.g. *Kickstarter*). Teenagers increasingly prefer communicating with their friends indirectly via texting, social media, and video-chatting rather than interacting face-to-face (Rideout and Robb, 2018). Furthermore, the proliferation of IT technologies has led to an upswing in "gig economy" and abundance of distributed enterprises. In these new forms of organizations, communications among team members are less often face-to-face than in the more traditional ones (see Marlow et al., 2017, for a recent review). Recently, the outbreak of COVID-19 and the resulting global lockdown have shown that a fast and sustainable digital transition in social and professional interactions is possible in a wide range of human activities.

Herein, we focus on the role of digital communication in building trust in human interactions. The main question we are asking in this paper is: *does the performance of a digital communication tool in inducing trust depend on its capacity to transmit nonverbal content?*[4] To answer this question, we conduct an incentivized laboratory experiment that implements a classic principal-agent paradigm (known as the hidden action game; see Charness and Dufwenberg, 2006) with one-way, pre-play communication. Before making decisions, the principal receives a message from the agent which is transmitted in one of the following ways: plain text message, auditory message, or video message. We believe that the three communication protocols studied in this paper are rather standard in the light of both the existing literature (see Bicchieri and Lev-On, 2007) and everyday experience. As we discuss in the next section, the novelty of our design allows us to homogenize the verbal content of communication across treatments, and thus isolate the causal

useful in conveying the particular information, and when that information is pertinent or appropriate to the social interaction" (p. 163). This echoes a broad definition of communication used in animal studies which we also adopt for the purpose of the present study: "behavior in [...] the sender [...] which evokes a response in [...] the receiver" (Ekman, 2006, p. 21). Clearly, this definition includes both verbal and nonverbal forms or components of communication.

[4]Numerous experiments point to the importance of nonverbal content of communication. A survey by Bonnefon et al. (2017) suggests that observing physical cues (such as facial displays or, more broady, bodily movements) may help detect cooperation in various economic games. See, for instance, Fetchenhauer et al. (2010); Bonnefon et al. (2013); Vogt et al. (2013); Centorrino et al. (2015); De Neys et al. (2015); van Leeuwen et al. (2017). Greiner et al. (2012) and Tognetti et al. (2020) provide related evidence for auditory cues. Furthermore, Eckel and Petrie (2011) and Zylbersztejn et al. (2020) document that people tend to exhibit a preference for receiving nonverbal cues in social interactions.

link between the richness of nonverbal content transmitted through a communication channel and the subsequent emergence of trust. We achieve this by inserting the recordings made in prior face-to-face interactions into our three communication treatments.

Our study aims to contribute to the experimental literature that highlights the importance of communication in fostering trust between humans and building sound economic partnerships. With some exceptions (that are discussed in the following section), it is commonplace in this research to employ a minimalistic communication protocol in which written messages are transmitted between anonymous parties. By design, these studies are well suited for investigating the role of verbal content, but not the nonverbal content. In the context of the hidden action game that we use, several experiments highlight a particular feature of verbal messages – a voluntary, non-binding promise to cooperate – as a credible signal of trustworthiness and an enhancer of trust.[5] This finding is also supported by field evidence from a TV show *The Golden Balls*, a high-stake prisoner's dilemma environment with a pre-play stage of natural face-to-face communication moderated by the host (Belot et al., 2010; Van den Assem et al., 2012; Turmunkh et al., 2019).

Our controlled laboratory setting allows us to extend the experimental testbed to a wide set of communication protocols – ranging from the minimalistic text messages to the most complex face-to-face interaction, and to focus on digital communication. For each of the studied environments, our experimental data corroborate the general finding that voluntary, non-binding promises promote trust. Furthermore, we extend this literature by shedding light on the role of nonverbal content of communication, as well as its interplay with the verbal content.

Our main findings are as follow. On aggregate, plain text messages engender less trust than either audio or video messages (which, in turn, yield similar trust rates). Providing auditory cues about the speaker seems to play a key role in inducing trust, as compared to providing a plain text content of the message. Adding visual cues in the video-based communication increases the observed trust rates only slightly (and not significantly). Furthermore, we also find evidence

[5] See, for instance, Charness and Dufwenberg (2006); Vanberg (2008); Ismayilov and Potters (2016); Schwartz et al. (2019). For a systematic review of the experimental literature on promise-making, see Woike and Kanngiesser (2019). Other characteristics of messages may also matter. Using a modified two-person trust game with an additional external observer, Chen and Houser (2017) report that other important features of a written message are its length (which increases trust) and the fact of mentioning money (which is associated with untrustworthiness). On the other hand, factors such as the use of encompassing words (like "we" or "us") do not seem to matter.

that these aggregate effects are moderated by the verbal content of communication. Including auditory cues generates a statistically significant increase in trust levels when the message conveys a promise to cooperate, with no further effect of visual cues. However, in the the absence of such a promise, a combination of auditive and visual cues is necessary to significantly increases trust levels compared to the baseline textual communication.

Finally, in the additional analyses we follow some of the previous experiments on strategic communication and focus on players' gender. We observe systematic gender differences in the environments with a restricted access to nonverbal cues. Females are generally less trusting than males when solely the verbal content of a message is transmitted (which usually does not even suffice to identify the other person's gender). Adding nonverbal content based on auditory cues helps close this gap. However, it also makes females display more trust towards other females than towards males (which is a bias, since it stands at odds with the degree of trustworthiness observed in both genders). These differences disappear once we further include visual cues about the sender.

Altogether, our experimental evidence points to the positive effect of the transmission of nonverbal content of communication on trust, but also to the heterogeneity of this effect. It may be sensitive not only to the factors that are endogenous to the process of communication (such as the verbal content that is transmitted by the sender alongside the nonverbal content), but also to the exogenous ones (such as players' gender).

2 Empirical strategy

The main objective of our study is to isolate the effect of the capacity to transmit nonverbal content on inducing trusting behavior in an experimental hidden action game with one-way, pre-play communication between players. The main virtues of our experimental design lie in the (i) ecological validity of the communication protocol (since we allow our subjects to address others in an unrestricted manner) while maintaining (ii) full control over the verbal content of communication across different communication modes, so as to (iii) draw causal inference about the impact of the

different layers of nonverbal communication on trust in economic interactions.

Our experiment relates to the early articles by Brosig et al. (2003) and Bicchieri and Lev-On (2007) who compare the performance of different modes of pre-play communication in inducing cooperation in public goods games. Like them, we employ a rich set of communication protocols: plain text message transmitted through a computer interface, audio message, and video message. The main findings from this literature are: (i) communication generally improves cooperation as compared to a no-communication control condition, (ii) communication via plain text messages and audio messages has a similar effect on cooperation, and (iii) communication via video messages and face-to-face communication perform equally well and further enhance cooperation.[6] However, such outcomes are far from being stable and the literature disagrees on the relative performance of various communication channels in inducing cooperative behavior. In a related experiment with public goods games, Bochet et al. (2006) find no difference between a communication condition with computerized plain text messages and another condition with face-to-face communication. The same null result is observed for trust and trustworthiness in a trust game experiment by Bicchieri et al. (2010).

Importantly, the determinants of behavior in this body of experiments are twofold: first, *how* people communicate (which is our variable of interest); second, *what* is the verbal content of their communication (which may impede the measurement of our variable of interest). Thus, these previous studies impose different communication channels as exogenous experimental treatments, but they leave the verbal content of messages to be endogenous and to vary freely across treatments.[7] Thus, behaviors observed in these studies may stem not only from an exogenous variation in the

[6] This echoes the findings from a meta-analysis by Balliet (2010). This study distinguishes between two forms of communication and reports that cooperation is higher under face-to-face communication than under communication via written messages. Other experiments provide similar evidence in the contexts of conflict resolution (Rockmann and Northcraft, 2008) and bargaining (Valley et al., 1998); see also Charness and Gneezy (2008) for a related discussion of earlier experimental evidence.

[7] Recent studies that investigate the relationship between the mode of communication and lying behavior (see, for example, Abeler et al., 2014; Conrads and Lotz, 2015; Cohn et al., 2018) do better in controlling the content of the communication across different channels. In these studies, subjects simply need to report the outcomes of coin tosses in all the communication channels (voice, text, online-form, or face-to-face). Abeler et al. (2014) find no difference in the reported outcome between phone call and on-line form. Cohn et al. (2018) report that the presence of a human counter part (who receives the reported outcome) is a more important determinant of lying than the mode of communication (text vs. voice). Finally, Naquin et al. (2010) find that people are more likely to lie when sending e-mails than in hand-written messages.

modes of communication allowing for different transmission of nonverbal content, but also from an endogenous and uncontrolled variation in the verbal content of communication. We conjecture that this may be the underlying cause of the diverging results in this literature. For the sake of illustration of the interplay between the treatment variable, the content of communication, and the eventual behavior, consider two examples drawn from the literature on promises.

Example 1. Using a set of communication conditions that is closely related to ours (spanning from computer chat to audio communication, then to face-to-face communication), Conrads and Reggiani (2017) report that the outcome of communication (i.e., the rate of promise-making) varies in a systematic way: the richer and more direct the communication protocol, the more likely a subject is to make a promise to the other person.[8] At the same time, they observe little variation in the subsequent rates of promise-keeping.

Example 2. In the context of the hidden action game, the data from Charness and Dufwenberg (2006, 2010) point to an opposite phenomenon: in a richer communication environment, subjects are less likely to make promises.[9] However, the effects of communication (overall, as well as conditional on a promise to be trustworthy) on trust and trustworthiness are stronger in the richer communication environment.

We also note that the existing literature proposes some methods for controlling specific features of communication. One way to control the variability of its content is to restrict the allowed topics of conversation (see Bouas and Komorita, 1996, for an early experiment applying this method). For instance, on top of varying modes of communication (either computer chat or face-to-face) in the trust game experiment by Bicchieri et al. (2010), players are either allowed or not allowed to make a promise. In a similar setting, Ismayilov and Potters (2016) allow pre-play written message which can or cannot be related to the experimental game. However, while controlling the

[8] Their experiment is based on an individual task rather than an experimental game. Communication takes place between a subject and the experimenter (a research assistant). In all communication conditions, the latter uses a standardized script presenting the task (an online survey to be filled in within 24h without any additional gratification) and asking a subject whether he or she promises to perform it.

[9] In those studies, authors consider two kinds of one-way, written messages: a fixed-form message (41/47 messages contain a promise to cooperate) and a free-form, hand-written statement (24/42 messages contain a promise to cooperate). We use Fisher's exact test to compare these two rates and obtain $p = 0.004$.

variation in the characteristic of interest, this method fails to control for the variation in other characteristics. For example, He et al. (2016) observe that when people are not allowed to talk about the game they are about to play, some of them may resort to using signs and body language to non-verbally transmit strategic content.

In this paper, we provide a simple yet novel experimental method for comparing the performance of different modes of digital communication, which solves the problem of the endogeneity of the verbal content of communication. We focus on one-way communication – from the agent to the principal – in order to capture the causal effect of the transmission of nonverbal content of communication on trust.[10] In order to homogenize the communication content across treatments, we first video-record messages transmitted in an experiment with direct (face-to-face, henceforth, F2F) interactions. Then, we use these recordings in our three main treatments. Therefore, all treatments are based on resampling from the same set of messages, and thus the verbal content is homogenous across treatments, while the richness of nonverbal cues transmitted in the communication process varies across communication modes.

Going beyond its purely methodological scope, the design of our experiment is also motivated by numerous real world situations – modern political, social, or commercial campaigns – in which messages conveyed to a relatively small group of people in a direct face-to-face manner are recorded, and then widely broadcasted via video-, audio-recording, or text transcripts. Importantly, the recent outbreak of COVID-19 and the resulting lockdown suggest that these forms of campaigning could become increasingly prominent in the future. As a dramatic example, consider the 2020 presidential campaign in Poland most of which happened under general lockdown. These unusual circumstance have greatly restricted live campaigning (e.g., election tours and partiasan rallys), leaving the candidates with little choice but to turn to small public events with large digital media broadcast. Our experimental design closely mimics such environments and can deepen our

[10] Anbarci et al. (2019) highlight the importance of two-way communication for enabling receivers to extract information that is both "precise" and "consistent in its implication for decision making" in a strategic context. We, however, note that when communication is real-time and multi-directional – like in Bochet et al. (2006), Bicchieri et al. (2010), several treatments in Brosig et al. (2003), as well as other experimental data included in Bicchieri and Lev-On (2007) and Balliet (2010) – each interacting group becomes an endogenously formed cell with no counterpart in other conditions. In this case, it becomes even harder (if not impossible) to establish a causal link from an individual message to behavior, since all the messages are correlated, and each player acts as both the sender and the receiver of messages.

understanding of the effectiveness of applying different communication techniques therein.

3 Experimental design and procedures

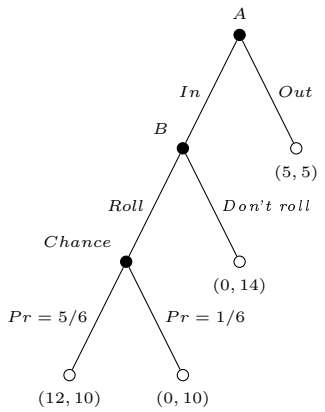
Our three main experimental treatments are based on a set of video recordings from the pre-play face-to-face communication from agents to principals gathered in our earlier experiment. We employ a classic hidden action game by Charness and Dufwenberg (2006) in our experiment. Participants acting as agents in those earlier sessions with face-to-face communication were clearly informed that they would be video-recorded and that these recordings could be used in future experiments. Participants to our three main treatments with digital communication are clearly informed that the recorded messages they are about to receive (either as video recordings, audio recordings, or text transcripts) come from the previous experimental sessions. Furthermore, the experimental instructions handed out to all participants specify that their own payoffs may depend on the earlier decision of the agent who sent the message, but that their own decisions have no further impact on that agent's payoffs.[11] Below, we outline the experimental game and the general procedure of our experiment. In the latter, we first describe the face-to-face experiment used to gather the recordings, and then our three main experimental treatments that exploit those aforementioned recordings.

3.1 Experimental hidden action game

Figure 1 presents the classic hidden action game by Charness and Dufwenberg (2006). All payoffs are in Euros. The game is played between two parties: the principal (henceforth, player A) and the agent (henceforth, player B). Player A may either choose an outside option *Out* which yields 5 to both players and ends the interaction, or go *In*. Then, player B may either choose to *Roll* a die (which yields 12 to A and 10 to B with the probability of 5/6, and 0 to A and 10 to B with the probability of 1/6), or not to *Roll* (yielding 0 to A and 14 to B with certainty). This game

[11] By the standard definition of trust adopted in the opening section – "the belief that others act in the interest of some measure of fairness or social welfare rather than their own self-interest", the way we link player A's payoff to player B's earlier decision allows us to measure player A's trust in player B's cooperative intentions even though both parties do not directly interact.

Figure 1: Experimental hidden action game



provides a simple setting for studying principal-agent relationships with moral hazard: incentives are not aligned between the two parties, and earning 0 is not perfectly informative for player A about player B's action.

Following Charness and Dufwenberg (2006), in the face-to-face implementation, we simultaneously elicit both players' decisions. Namely, the player B makes a decision without knowing player A's decision. Player B's decision is only implemented had player A gone *In*. In our three main treatments the player B's decision is pre-recorded, and again player A takes the decision without observing it. In addition, in all experimental conditions the game is preceded by a pre-play communication stage in which player B delivers a message to player A.

3.2 Procedure

First, we have collected the dataset for the direct face-to-face (F2F) communication condition. This auxiliary condition plays an instrumental role in the main experiment: the preliminary dataset collected therein allows us to implement the three digital communication conditions. We have video-recorded all the pre-play statements made by player Bs, and kept records of their subsequent choices (to *Roll* or not), as well as of the outcomes of die rolls. Based on that information, we have then constructed our three treatments of interest – VIDEO, AUDIO, and TEXT – in which we vary the channel through which player Bs' messages are transmitted to player As.

Experiment begins with preliminary measurements (distributional preferences, risk preferences, cognitive skills, emotional intelligence, general trust towards others) and a basic socio-demographic questionnaire. This is followed by treatment-specific experimental instructions. In VIDEO, AUDIO, and TEXT, those instructions also include a one-page summary of the F2F instructions. Furthermore, after making the decision in the game (to go *In* or to stay *Out*), player As fill in a short questionnaire about their perceptions of player Bs.[12] Below, we overview the implementation of this hidden action game.

3.2.1 F2F

Each experimental sessions involves 6 player As and 6 player Bs. All player As remain in one room during the whole experiment. They are seated in a single row, isolated one from another by separators, and not allowed to talk. The space in front of them is left open and used by a player B to make a brief statement. A video camera recording player Bs' statement is discretely positioned in the middle of the player A's raw.

Player Bs enter the room one by one, so that player As play six rounds of the game (which is common knowledge). Each time, player B faces the center of player As' row, and all player As have a clear view on the speaker. Player B also has a clear, unobstructed view on all six player As.

After making a statement, player B is invited to a separate room where s/he privately decides whether to *Roll* a die or not. Then, s/he is asked to leave the laboratory and wait outside until the end of the experiment. At the same time, each player A makes a decision whether to go *In* or stay *Out*. All decisions are made on a sheet of paper, which is then put in an envelope, sealed, and collected by the laboratory staff after each round. In addition, once player B has made a decision and left the separate room, a laboratory staff member rolls a die in private and marks the outcome on player B's sealed envelope.

[12] Note that the experimental instructions use different labels for player A's decisions than the one used by Charness and Dufwenberg (2006) – *Right* instead of *In* and *Left* instead of *Out*. In the paper we adhere to their original nomenclature. An English translations of the instructions, a summary of the measured characteristics, and further details on how the experiment was implemented in the lab (including extra care we have put in to minimize the likelihood that player As and player Bs know each other) are provided in Appendices A, B, and C, respectively.

At the end of the experiment, player As and player Bs are randomly and anonymously matched in pairs. The outcome of the game for each pair is based on the payoff structure described in Figure 1 and defined by the decision made by player A after player B's statement, as well as the decision made by player B in a private room had A chosen to go *In*. For B's decision to *Roll*, the outcome of the die roll is also taken into account.

3.2.2 VIDEO, AUDIO, and TEXT

In these three treatments, each participant acts as a player A and is exposed to 10 items randomly drawn from the set of 41 recordings of player Bs' statements made in F2F.[13] Subjects are clearly informed that these recordings have been gathered in experiments conducted in the past, and that although their own decisions do not influence the payoffs of the player Bs behind those messages, their own payoffs may depend on those player Bs' antecedent decisions.

In each round, each participant receives a message (a statement made by player B). In the VIDEO treatment, participants watch the video recording of the message. In AUDIO, they listen to an audio recording. They can watch/listen to a given recording only once. In TEXT, precise transcripts of the statements are displayed on player As' computer screens for the amount of time equal to the duration of the corresponding recording.[14] We would like to emphasize that the TEXT condition is not meant to measure the potential of text messages, or more generally written statements, in inducing trust in humans. Rather, TEXT is the baseline condition through which we control for the effect of merely providing verbal content of communication on subsequent trust, so as to isolate the effects of nonverbal content. Hence, the transcripts keep track of vocal parasites, slips of the tongue, sudden pauses, repetitions, etc., which are common for improvised spoken language and come as an essential feature of what we consider as the verbal content of communication (which, in turn, is meant to remain constant across treatments).[15]

[13] We have conducted 7 F2F sessions. However, one player B in session 6 of F2F decided to quit after the preliminary measurements and before receiving the instruction of the hidden action game, and was replaced by a research assistant unknown to player As. To avoid any contamination of player As' behavior, that research assistant acted as player B in the final round of the experimental game. The data from that round were dismissed and our dataset from that session only covers 5 player Bs, and thus 41 player Bs in total.

[14] Pilot tests suggested that this exposure time was sufficient for reading with comprehension a short statement displayed on a computer screen.

[15] Christenfeld (1995) shows that "ums" may matter for the receiver's perception of the sender. However, we also

After receiving a message, each participant decides whether to go *In* or stay *Out* and fills in the post-decision questionnaire. Participants are not informed about the choice made by the player B whose message they have just received, or about the outcome of die roll.

After completing 10 rounds, each player A is randomly matched with one of the ten player Bs. The payoff structure corresponds to player A's payoffs in Figure 1. Player A receives 5 Euros when staying *Out*, 0 Euros when going *In* with a player B who does not *Roll*, or either 12 Euros (with 5/6 chance) or 0 Euros (with 1/6 chance) if player B chooses to *Roll*.

The experiment took place at Laboratoire d'Economie Expérimentale de Nice (LEEN) of the Université Côte d'Azur, France, between April 2017 and February 2018.[16] All treatments were implemented through a between-subject design. In addition to 83 subjects who participated in F2F experiment, a total of 217 subjects participated in VIDEO (62), AUDIO (70), and TEXT (85) treatments.[17] Our subjects are predominantly students (296/300), their average age is 21.28 (SD 3.69), 59.3% of them are females. See Appendix D for a summary of other characteristics of our sample.

4 Results

4.1 Summary of player Bs' behavior

Before investigating player As' behavior, let us first briefly summarize the behavior observed among their counterparts. 61% of player Bs in the F2F experiment decide to *Roll*, while a promise to *Roll* is made by 51.2% of them. Following Charness and Dufwenberg (2006), a promise is defined as statement of intent to *Roll*.[18] There is no significant difference between the duration of the

believe that the mere presence of inarticulate sounds does not affect the interpretation of the strategic elements in the verbal content of messages. Consistently with a large body of experiments using more "polished" free-form, hand-written messages, in our TEXT condition with raw transcripts we identify a positive effect of promise-making on trust (as discussed in the following section).

[16] The computerized parts of the experiment are all programmed in z-tree (Fischbacher, 2007).

[17] Number of participants in VIDEO, AUDIO, and TEXT varies due to the natural variation in show-up rates across sessions. The recruitment of the participants has been carried out using ORSEE (Greiner, 2015).

[18] As raised by Houser and Xiao (2011), the *ex post* interpretation of free-form messages is a major methodological challenge for the experimenter. The literature still lacks a common consensus on whether this should involve content analysis carried out by the experimenter (Charness and Dufwenberg, 2006), by independent coders (He et al., 2016), through an incentivized coordination game (Houser and Xiao, 2011), or by asking the subjects for their own interpretation (Servátka et al., 2011). Echoing a recent study by Schwartz et al. (2019), herein we implement a

messages with promise as compared to the ones without (average duration: 28.90 seconds vs. 23.75 seconds; $p = 0.108$ using ranksum Mann-Whitney test).

There is a large difference between the frequencies of *Roll* conditional on a promise to *Roll* (71%) and without such promise (50%). The result of a logit regression of the *Roll* dummy on the *Promise to Roll* dummy (and session fixed effects) suggests that a promise to *Roll* is highly predictive of a subsequent decision to *Roll*: the average marginal change of the likelihood of *Roll* due to a *Promise to Roll* is estimated at 0.394 (SE=0.122, $p = 0.001$).[19] In the additional analyses reported in Appendix F, we also offer a screening exercise exploiting the information about player B's individual characteristics to investigate who makes a promise. We find that pro-social preferences (measured by the SVO angle) stand out as a significant predictor of promise-making.

The *Roll* (promise-making) rates in our three treatments are as follow: 66.6% (53.9%) in VIDEO, 64.7% (52.9%) in AUDIO, and 60.0% (49.3%) in TEXT. These small variations are due to random sampling, and both rates are equally distributed across treatments.[20] Albeit small, these sampling differences are nonetheless accounted for in the comparisons of aggregate trust rates across treatments (see Section 4.2). First, we check if there are important differences in aggregate trust rates for those player Bs that chose to *Roll* and those who did not. These differences are found to be small (0.4, 2, and 3 percentage point in VIDEO, AUDIO, and TEXT, respectively) suggesting that the observed treatment effects cannot simply stem from those random sampling variations in player Bs' trustworthiness. Second, echoing the previous experimental studies of communication and trust, we condition the trust rates on the content of messages (i.e., promise vs. no promise to *Roll*).

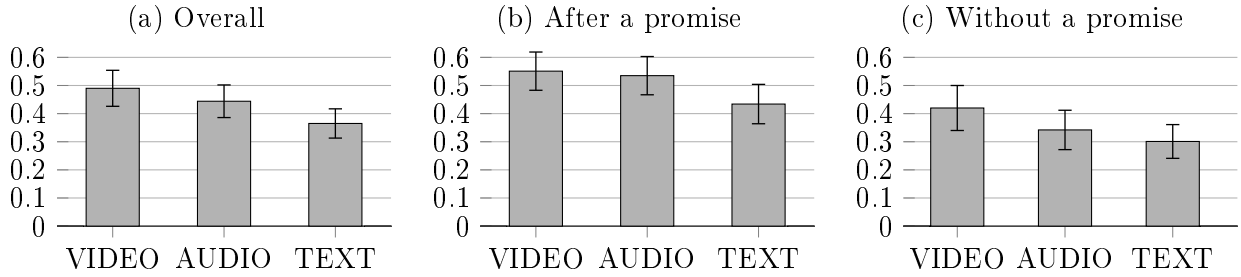
---

method combining the first two approaches. Player Bs' statements were classified as promises or non-promises by two independent coders (research assistants). The first coder classified the content of messages (promise or not) while preparing the transcripts of player Bs' statements. Then, another coder received the full list of 41 transcripts and independently classified each of them as a promise or a non-promise. Ties were broken by one of the authors. All the transcripts (translated from French to English) along with their classification (promise or non-promise) and the resulting decision (*Roll* or *Don't roll*) may be found in Appendix E.

[19] Session dummies (7 sessions, first session dummy dropped) allow us to account for the fact that player Bs face the same audience of player As within a session, but the audiences vary between sessions. These coefficients are jointly insignificant ($p = 0.287$). For details, see Table A.3 in Appendix F.

[20] To test whether each of these variables is equally distributed across the three treatments, for every player A we calculate the frequency of decision *Roll* and the frequency of promise-making among the ten player Bs that person faced during the experiment. Comparing the distributions of those frequencies across the three treatments using Fisher's exact test, we find $p = 0.258$ for *Roll* and $p = 0.123$ for the promise to *Roll*. Given that the size of each contingency table is relatively large (9×3), both p-values are computed using Monte Carlo simulations.

Figure 2: Average Trust Rate



Note. Each bar depicts the aggregate likelihood of a player As going *In*: across treatments (Panel (a)), and conditional on the type of player B's statement (Panels (b) and (c)). The error bar shows two standard errors range.

4.2 Trust and promises with digital communication

Figure 2 summarizes aggregate trust rates across our experimental conditions. The statistical comparisons of these rates are based on linear probability models (LPM) reported in Table 1. In those models, we regress the trust indicator variable (= 1 if player A goes *In*, = 0 otherwise) on the set of condition indicator variables corresponding to various conditions (= 1 for a given condition, = 0 otherwise). The models are constructed such that each coefficient corresponds to the trust rate observed in a given condition. We estimate cluster-robust standard errors (individual-level clustering) to account for the potential unobserved correlation between individual observations and the heteroscedasticity of residuals in the LPM.

The first model (Table 1a) uses treatment indicators as a set of explanatory variables. In the second model (Table 1c), these variables are further interacted with promise/non-promise indicators. For each model, the equality of trust rates between two (multiple) experimental conditions is parametrically tested by comparing different coefficients through t-test (F-test). Finally, in Tables 1b and Table 1d we compare aggregate trust rates across the experimental conditions.[21]

Panel (a) of Figure 2 provides the average trust rates, i.e. the likelihood that a player A goes *In*, in VIDEO, AUDIO, and TEXT.[22] Overall, we observe similar trust rates in VIDEO and AUDIO (49.0% and 44.4%, respectively; $p = 0.289$) which are, in turn, significantly higher than

[21] As a robustness check, Table A.5 in the Appendix G re-assesses the aggregate comparisons from Table 1b after including player B fixed effects. This is a general parametric way to account for the fact that different player As are exposed to a message coming from the same player B. All the results remain robust to this extension.

[22] See Appendix H for details (as well as caveats) of the comparisons of these three treatments with the F2F experiment. In a nutshell, the trust rate in F2F is 40.2% – descriptively higher than in TEXT, and lower than in AUDIO and VIDEO; yet, none of those differences is found to be statistically significant.

Table 1: Statistical support for Figure 2 and trust rate comparisons

	Rate (SE)	p
$1[VIDEO]$ (α_1)	0.490	< 0.001
	(0.032)	
$1[AUDIO]$ (α_2)	0.444	< 0.001
	(0.029)	
$1[TEXT]$ (α_3)	0.365	< 0.001
	(0.026)	

(a) Trust rates and their SE from Fig. 2a

	Diff. (SE)	p
VIDEO vs. AUDIO	0.046	0.289
($H_0 : \alpha_1 - \alpha_2 = 0$)	(0.043)	
VIDEO vs. TEXT	0.126	0.003
($H_0 : \alpha_1 - \alpha_3 = 0$)	(0.042)	
AUDIO vs. TEXT	0.080	0.042
($H_0 : \alpha_2 - \alpha_3 = 0$)	(0.039)	

(b) Comparing trust rates from Fig. 2a

	Rate (SE)	p
$1[VIDEO, NON-PROMISE]$ (β_1)	0.420	< 0.001
	(0.040)	
$1[AUDIO, NON-PROMISE]$ (β_2)	0.342	< 0.001
	(0.035)	
$1[TEXT, NON-PROMISE]$ (β_3)	0.301	< 0.001
	(0.030)	
$1[VIDEO, PROMISE]$ (β_4)	0.551	< 0.001
	(0.034)	
$1[AUDIO, PROMISE]$ (β_5)	0.535	< 0.001
	(0.034)	
$1[TEXT, PROMISE]$ (β_6)	0.434	< 0.001
	(0.035)	

(c) Trust rates and their SE from Fig. 2b and 2c

	Diff. (SE)	p
Trust rates w/o a promise to *Roll*		
VIDEO vs. AUDIO	0.077	0.151
($H_0 : \beta_1 - \beta_2 = 0$)	(0.054)	
VIDEO vs. TEXT	0.119	0.020
($H_0 : \beta_1 - \beta_3 = 0$)	(0.050)	
AUDIO vs. TEXT	0.042	0.371
($H_0 : \beta_2 - \beta_3 = 0$)	(0.046)	
Trust rates after a promise to *Roll*		
VIDEO vs. AUDIO	0.016	0.742
($H_0 : \beta_4 - \beta_5 = 0$)	(0.048)	
VIDEO vs. TEXT	0.117	0.016
($H_0 : \beta_4 - \beta_6 = 0$)	(0.048)	
AUDIO vs. TEXT	0.101	0.037
($H_0 : \beta_5 - \beta_6 = 0$)	(0.048)	
The effect of promise on trust		
VIDEO	0.131	0.000
($H_0 : \beta_4 - \beta_1 = 0$)	(0.036)	
AUDIO	0.193	0.000
($H_0 : \beta_5 - \beta_2 = 0$)	(0.038)	
TEXT	0.133	0.000
($H_0 : \beta_6 - \beta_3 = 0$)	(0.037)	
Joint equality of diff.'s		0.420

(d) Comparing trust rates from Fig. 2b and 2c

Note. *(a) and (c)*: OLS estimates from linear probability models (LPM) regressing the trust indicator variable (= 1 if player A goes *In*, = 0 otherwise) on the all-experiment set of condition indicator variables (= 1 for a given condition, = 0 otherwise). The models are intercept-free, and each coefficient provides the aggregate rate ("Rate") of trust in a given experimental condition. p-values correspond to standard t-test of a coefficient's nullity.
(b) and (d): coefficient comparisons based on the models (a) and (c), respectively. p-values correspond to t-tests for the nullity of the differences ("Diff.") between a given pair of coefficients; the final test in (d) is an F-test for the joint nullity of three differences.
In all models, observations are clustered and the individual level and estimated standard errors are cluster-robust. There are 620 observations (62 clusters) in VIDEO, 700 (70) in AUDIO, and 850 (85) in TEXT.

in TEXT (36.5%, with $p = 0.003$ vs. VIDEO, and $p = 0.042$ vs. AUDIO).[23]

At the treatment level, these rates are insensitive to the conditioning on player B's actual trustworthiness: *ex post*, we hardly observe any variation between the aggregate rates of trust displayed towards player Bs who decided to *Roll* as compared to those who did not (VIDEO: 49.2% vs. 48.8%; AUDIO: 45.5% vs. 42.5%; TEXT: 37.3% vs. 35.3%). This suggests that, in aggregate terms, player As' trust is disconnected from player Bs' trustworthiness. However, a message may nonetheless convey a credible signal of player B's trustworthiness that is taken into account by player As: a voluntary promise to *Roll*.[24]

Panels (b) and (c) of Figure 2 separates the trust rates depending on whether player B's message contains a promise to *Roll* (Panel b) or not (Panel c). Trust rates following a promise do not significantly vary between VIDEO and AUDIO (55% vs. 54%, $p = 0.742$), and they are both significantly higher than 43% observed in TEXT ($p = 0.016$ vs. VIDEO, $p = 0.037$ vs. AUDIO). Without a promise, trust rates vary from 42% in VIDEO to 34% in AUDIO to 30% in TEXT, and this time only the difference between VIDEO and TEXT is significant ($p = 0.020$; other pairwise comparisons yield $p > 0.150$). The effect of a player B's promise to *Roll* on the overall trust rate is positive (13 percentage points in VIDEO and TEXT, and 19 percentage points in AUDIO) and statistically significant in all the treatments ($p < 0.001$). Furthermore, this effect does not vary significantly across the three treatments ($p = 0.420$, F-test).

To further investigate the robustness of the effect of player Bs' promises on player As' trust, Table 2 provides average marginal effects from a logistic regression of the *In* dummy on the $1[Promise]$ dummy, coupled with a set of individual-level control variables. These controls are related to two questions: *who* is the subject in the role of player A that is more likely to trust, and *how* does that person perceive the participant in the role of player B?

[23] Another way to look at those data is to take each player B as an individual observation and compare the trust rates (i.e., the likelihood that a given player B is trusted by a player A) across the three experimental conditions using two-sided signrank test. The main patterns in the data remain unchanged: trust rates are lower in TEXT than in either AUDIO ($p = 0.012$) or VIDEO ($p = 0.001$), and do not differ significantly between the last two conditions ($p = 0.241$). The respective comparisons conditional on a promise (on a non-promise) yield $p = 0.013$, $p = 0.005$, and $p = 0.715$ ($p = 0.390$, $p = 0.100$, and $p = 0.204$).

[24] This observation echoes Belot et al. (2010) and Van den Assem et al. (2012) who use prisoner's dilemma data from a high-stake TV with a pre-play, free-form communication stage (with moderation from the host), and find (*i*) no evidence of a correlation in players' decisions on aggregate, but also (*ii*) point to the importance of voluntary promises in inducing individual cooperative behavior.

Table 2: Promises and individual predictors of trust: marginal effects from logit regression models

	VIDEO		AUDIO		TEXT	
	$ME, (SE)$	p	$ME, (SE)$	p	$ME, (SE)$	p
$1[Promise]$	0.103	0.001	0.175	< 0.001	0.100	0.002
	(0.030)		(0.039)		(0.032)	
Player A's perception of player B's characteristics (1-8 scale):						
beauty	-0.016	0.232	-0.010	0.449		
	(0.014)		(0.013)			
intelligence	0.033	0.070	0.019	0.352	0.066	< 0.001
	(0.018)		(0.020)		(0.010)	
trustworthiness	0.088	< 0.001	0.058	< 0.001	0.033	< 0.001
	(0.013)		(0.015)		(0.009)	
self-confidence	-0.009	0.465	-0.001	0.920	-0.007	0.461
	(0.012)		(0.012)		(0.009)	
Player A's characteristics:						
SVO angle	-0.002	0.249	< 0.001	0.928	< -0.001	0.954
	(0.002)		(0.002)		(0.002)	
$1[CRT > 0]$	0.020	0.759	0.070	0.238	-0.052	0.431
	(0.066)		(0.059)		(0.066)	
RMET score	0.003	0.680	-0.004	0.469	0.002	0.782
	(0.007)		(0.005)		(0.006)	
GP investment	0.001	0.285	0.003	0.013	< 0.001	0.988
	(0.001)		(0.001)		(0.001)	
SOEP trust 1	-0.074	0.161	-0.072	0.172	-0.068	0.207
	(0.053)		(0.053)		(0.054)	
SOEP trust 2	-0.003	0.953	0.097	0.006	0.084	0.072
	(0.050)		(0.035)		(0.046)	
SOEP trust 3	0.081	0.260	-0.016	0.813	0.116	0.118
	(0.072)		(0.068)		(0.074)	
Political views	-0.001	0.943	0.003	0.840	0.003	0.767
	(0.017)		(0.013)		(0.010)	
Pseudo-R^2	0.142		0.100		0.115	
Nb of obs./clusters	620/62		700/70		850/85	

Note. Average marginal effects from logistic regression models. First/second/third model: data from VIDEO/AUDIO/TEXT treatment. Observations are clustered and the individual level and estimated standard errors are cluster-robust.

To empirically address the first question, we exploit the set of individual characteristics we have gathered. For the second question, we exploit the information about player A's perceptions of player B's characteristics: beauty (collected everywhere but in TEXT), intelligence, trustworthiness, and self-confidence, all measured on a 1-8 scale.[25] Since this information has been elicited from every player A in each round only after the decision-making stage of the game, we stress the correlational nature of its relationship with the dependent variable.[26]

The main result of this exercise is that the effect of a promise on trust remains positive and significant even after controlling for a wide range of individual-level variables. Not surprisingly, we also find that player As are more likely to trust player Bs the more they see them as trustworthy. We also observe a positive association between one's trust and the perceived intelligence of the other person in most conditions ($p = 0.070$ in VIDEO; $p < 0.001$ in TEXT; insignificant in AUDIO). The perceived beauty or self-confidence of a player B are not significantly predictive of a player A's trust. Finally, in the absence of visual cues about player Bs, player As' behavior seems to be guided by their general sense of trust towards strangers: the marginal effect of "SOEP trust 2" (i.e., the answer to the question "How much do you trust people you just met?" on a 1-4 scale; see Appendix B.1) is positive and significant at the 1% level in AUDIO ($p = 0.006$) and at the 10% level in TEXT ($p = 0.072$) models, but not in the VIDEO model ($p = 0.953$).

[25] Although promise-makers receive slightly higher average scores from player As for each of the four characteristics (see Appendix I), we fail to detect any significant differences between promise and perceived trustworthiness across treatments. For each condition, ranksum test using player Bs' average score on trustworthiness as independent observation fails to identify significant differences between promise-makers ($N = 21$) and others ($N = 20$). Regarding player Bs' intelligence, the same test indicates significant differences in VIDEO and AUDIO.

[26] We also note that the previous evidence suggests that at least part of this relationship may be causal. In a related experiment, Greiner et al. (2012), introduce a three-person dictator game in which each of the two recipients may independetly engage in a one-way communication with the decision-maker. They allow for different pre-play communication protocols: either none or through (pre-recorded) muted video messages for both recipients or through a muted video message for one recipient and a video message with sound for the other. Like us, they elicit the decision-makers' internal (i.e., post-decision) ratings of their interaction partners individual traits and compare them with those provided independently by external rates. The conclude that "at least part of the internal ratings are based on a general, unbiased, common impression, independent of previously made decisions", so that choices are to a large extent driven by social evaluations, although "also idiosyncratic social tastes play a role, which [...] are indistinguishable from ex-post rationalization of previous [...] decisions" (pages 411-412; see their Result 4).

4.3 Additional analyses: gender differences in trust

In the previous section, we have shown how the role of the different layers of nonverbal communication depends on a factor that is endogenous to the communication process, namely the verbal content of statements, while also accounting for the receiver's subjective perception of the sender. Herein, we exploit the natural variation of players' gender in our dataset to illustrate the possible sensitivity of the effect of the nonverbal content of communication on trust to a characteristic that is exogenous to the communication process. We see this exercise as an extension of the related experimental study by Greiner et al. (2012) based on a sample of female subjects.

We note that the existing experimental literature on communication lacks a consensus on the relationship between gender and trust. Belot et al. (2010, 2012) and Van den Assem et al. (2012) analyze a dataset from a high-stake prisoners' dilemma game played in a TV show that females are more likely to share the prize than males, and that external observers can accurately predict this gender difference. Bicchieri et al. (2010), in turn, find no statistical evidence for gender differences in trust and trustworthiness in a trust game played with different modes of communication (either computer-mediated or face-to-face), and disregard this aspect in their analyses; similar conclusions can be found in Chen and Houser (2017).

In our data, gender does not appear to be an important predictor of player Bs' behavior. The promise-making rates are fairly similar for both genders (females: 9/20; males: 12/21; Fisher's exact test yields $p = 0.538$). The average marginal effect of being a female on the likelihood of choosing to *Roll* is negative (-0.113), yet insignificant ($p = 0.479$).[27] Thus, we interpret any difference in player As' behavior that stems from player Bs' gender as a behavioral bias.

For the sake of empirical analysis, we suppose that people may have different trust attitudes towards males and females, depending on their own gender (which is known to the experimenter) and some other (unobserved) preferences. To capture these individual trust attitudes, for each player A, we construct two variables: *Ratio trust M* and *Ratio trust F*. They capture the fraction of decisions *In* in all the interactions with a player B of either gender (M stands for male, F stands

[27] This result stems from a logit regression with session fixed effects, and holds (marginal effect of -0.088, $p = 0.534$) once we additionally control for the presence of a promise to *Roll* (the marginal effect of which is once again positive, 0.390, and highly significant, $p = 0.001$).

Table 3: Trust and gender: aggregate results

	VIDEO		AUDIO		TEXT	
Ratio trust	*F*	*M*	*F*	*M*	*F*	*M*
Male player A	54.1%	49.6%	53.5%	46.4%	43.0%	43.7%
Female player A	48.9%	46.2%	50.4%	37.7%	30.7%	32.4%

Note. For each player A, *Ratio trust M* (*F*) provides the individual rate at with that player trusted male (female) player Bs. The average individual rates in the sample are conditioned on treatment and player A's gender.

for female).[28] The sample means of those individual ratios are summarized in Table 3.

We then conduct the following estimation exercise. For ith player A, the overall trust ratio towards player Bs of gender $g \in \{F, M\}$ depends on i's gender in the following way:

$$Ratio\_trust_{ig} = \gamma_0 + \gamma_1 \times 1[Female\_A_i] + u_{ig}, \tag{1}$$

where $1[Female\_A_i] = 1$ if an ith player A is a female, and $= 0$ otherwise. Since both $Ratio\_trust_{iF}$ and $Ratio\_trust_{iM}$ may also stem from i's unobserved preferences (so that u_{iF} and u_{iM} may be correlated), these models are estimated as seemingly unrelated regressions (see, Cameron and Trivedi, 2005, pp. 209-210) and presented in Table 4. A notable difference between TEXT and the remaining treatments is that in the former, player B's gender remains mostly unknown to player As, so that $Ratio\_trust_{iF}$ and $Ratio\_trust_{iM}$ in that treatment should be similar.[29]

The main results are as follows. In TEXT (where player B's gender is predominantly unobservable for player As), female player As are found to be generally less trusting than male player As ($H_0 : \gamma_1 = 0$ for *Ratio trust F*: $p = 0.041$; for *Ratio trust M*: $p = 0.055$). For both male and female player As, we do not find a significant difference between the two rates: coefficients γ_0 as well as $\gamma_0 + \gamma_1$ are not different in equations *Ratio trust F* and *Ratio trust M* ($p = 0.871$ for γ_0 and $p = 0.629$ for $\gamma_0 + \gamma_1$). In AUDIO, male player As' trust does not vary as a function of player

[28] On average, a player A encounters a female player B in 4.71 rounds out of 10 in VIDEO, in 4.57 in AUDIO, and in 5.08 in TEXT. Replicating the procedure described in footnote 20 for these data gives $p = 0.287$. Appendix C provides further details on the gender composition of the experimental sample.

[29] Only 3 out of 41 transcripts reveal the sender's gender: twice as a female and once as a male. In French, the grammar rule of *accord* may impose gender-specific forms of verbs (depending on the tense) and adjectives in written which, however, are not always distinguishable in spoken (so that the masculine and the feminine forms may sound exactly the same). In our transcripts, we respect the French grammar and follow the rule of *accord*.

Table 4: Trust and gender: seemingly unrelated regressions

	VIDEO		AUDIO		TEXT	
Ratio trust	*F*	*M*	*F*	*M*	*F*	*M*
Intercept (γ_0)	0.541^a	0.496^a	0.535^a	0.464^a	0.430^a	0.437^a
	(0.065)	(0.061)	(0.067)	(0.056)	(0.046)	(0.045)
$1[Female\_A]$ (γ_1)	-0.052	-0.034	-0.031	-0.087	-0.123^b	-0.113^c
	(0.080)	(0.076)	(0.081)	(0.069)	(0.060)	(0.059)

Note. For each treatment, we provide estimated coefficients (and standard errors) of seemingly unrelated regression models. The dependent variables are *Ratio trust F* and *Ratio trust M*, the explanatory variable is an indicator variable $1[Female\_A]$. $a/b/c$ indicate statistical significance at the 1/5/10% level.

B's gender (coefficients γ_0 do not vary between the two equations, $p = 0.283$). However, female player As now exhibit stronger trust towards player Bs of their own gender: the model suggests that $\gamma_0 + \gamma_1$ is statistically significantly higher in the equation *Ratio trust F* than in *Ratio trust M* ($p = 0.007$). On the other hand, we also note that the effect of player A's gender on trust in AUDIO is less pronounced than in TEXT: for neither ratio, we detect a statistically significant difference between male and female player As ($H_0 : \gamma_1 = 0$ in *Ratio trust F*: $p = 0.701$; *Ratio trust M*: $p = 0.208$). Finally, all differences between trust ratios fade away in the VIDEO treatment.

To conclude, we find evidence for gender differences in trust that can be overridden by means of providing nonverbal content in the process of communication. First, when no nonverbal content is transmitted and player B's gender remains unknown (TEXT treatment), we observe that female player As are generally less trustful than males. The AUDIO treatment (in which the auditory cues provide straightforward information about the speaker's gender) helps close this gap. However, it also gives rise to a gender bias in trust: female player As tend to favor players Bs of their own gender. Finally, any gender differences fade away in VIDEO that enriches the nonverbal content by visual cues about the sender.

5 Discussion

A large body of experimental research in economics and other social sciences explores the role of communication in various types of economic interactions, such as cooperation, coordination,

trust and reciprocity (see, among others, Ostrom et al., 1992; Cooper et al., 1992; Charness, 2000; Duffy and Feltovich, 2002; Ellingsen and Johannesson, 2004; Duffy and Feltovich, 2006; Bochet and Putterman, 2009; Bracht and Feltovich, 2009; Gangadharan et al., 2017). This body of empirical evidence generally points to the conclusion that communication increases the efficiency of interactions (Sally, 1995; Crawford, 1998; Balliet, 2010). Here, we complement this literature by asking a different question: how important is the nonverbal content of communication in inducing trust? Overall, our experimental data highlight an interplay between verbal and nonverbal content of communication in the process of inducing trust in human interactions.

To answer our research question, we use a controlled laboratory experiment based on a principal-agent relationship with different modes of one-way communication. Our design allows us to homogenize the verbal content of communication across the experimental conditions. We document that the degree to which communication channels enable the transmission of nonverbal content is important for the emergence of trust in human interactions. In aggregate terms, providing auditory cues about the speaker seems to play a key role in inducing trust, as compared to providing a plain text content of the message. Adding visual cues in video-based communication does not further lead to a statistically significant increase in the observed rates of trust, notwithstanding the general findings in Bicchieri and Lev-On (2007). However, we find evidence that these effects are moderated by the verbal content of communication. They hold for messages containing a promise to cooperate (which constitutes a significant and well recognized predictor of trustworthiness); in the absence of such a promise, a combination of auditive and visual cues is necessary to significantly enhance trust as compared to a plain text message.

These results complement and extend the previous evidence reported in Greiner et al. (2012). Using three-person dictator game in which each of the two recipients may independently engage in one-way communication with the decision-maker, they allow for different pre-play communication protocols: either none or through (pre-recorded) muted video message from each recipient, or through a muted video message for one recipient and a video message with sound for the other. While allowing for identification of the recipient via silent video does not affect the average levels of donations as compared to the baseline condition, average donations increase when the auditory

channel is added. This result is consistent with our data in that it points to the importance (in aggregate terms) of auditory cues relative to visual cues.

Our study extends theirs in following ways. Unlike their dictator game, our hidden action game involves strategic interactions. Their data do not reveal any impact of the content of the recipients' messages on the decision-makers' behavior, which may be due to the fact that the conveyed information necessarily lacks strategic importance. Our evidence suggests that when messages can convey strategic signals (e.g., related to promise-making), the content of communication may affect other-regarding behavior alongside the channel of communication.

Furthermore, Greiner et al. (2012) report that the decision-makers in their experiment form impressions about the recipients and discriminate them on this basis. However, since their experiment employs a sample of female participants, an open question is how discrimination in other-regarding behavior may depend on both players' gender.

In our additional analyses, we explore whether and how digital communication protocols may give rise to gender differences in trust. The most restricted communication environment in which players send plain text messages and that makes it virtually impossible to identify the other person (not even his or her gender), gives rise to a gender difference in trust: females tend to be generally less trustful than males.[30] Providing auditory cues that reveal the sender's gender leads to a gender bias in trust: female principals tend to trust other females more than to trust other males, even though gender is not predictive of the agents' trustworthiness in our experiment (if anything, females happen to be slightly, yet insignificantly, less trustworthy). All those differences disappear once visual cues also become available.

We see two potential avenues for future research. First, economic models of strategic communication in cooperation problems focus on the content of messages. In particular, behavioral theories stipulate that communication may help building commitment to cooperate through promise-making (see, e.g., Ellingsen and Johannesson, 2004; Charness and Dufwenberg, 2006). Combined with the aforementioned experimental studies, our experimental evidence suggests that future theoretical developments should also account for the channel through which messages are

[30] This stands in line with Buchan et al. (2008) who report that men are more trusting than women. However, they also find that women are more trustworthy which is not confirmed by our data.

transmitted. Second, despite its methodological virtues, our study has one limitation that is worth emphasizing in the closing lines. The stimuli used herein is based on a specific sample of 41 recordings that has been resampled and employed throughout the whole experiment. Thus, an improved experimental control and the ability to draw causal inference about the effect of the digital communication mode on behavior comes with a price of reducing the natural heterogeneity of human communication (related to the variation in the content of messages, as well to the variety of senders' individual characteristics). Thus, there is certainly a need for further evidence, as well as out-of-the-sample replications.

Acknowledgments

This is a revised version of GATE WP 1909 entitled "Digital communication and swift trust". Ynès Bouamoud, Imen Bouhlel, Maxime David, Carla Lapointe, Ismaël Rafaï, Charlotte Saucet, Ann Thampi, Vincent Théroude, Mira Toumi, Dorian Vincileoni, and Charlène Wojerz have provided quality research assistance. Quentin Thévenet has provided valuable assistance with software programming. We thank the editor Anthony Kwasnica and two anonymous referees, as well as Giuseppe Attanasi, Brice Corgnet, John Duffy, Martin Dufwenberg, Astrid Hopfensitz, Nicolas Jacquemet, Annett John, Michal Krawczyk, Charles Noussair, Ro'i Zultan, Diego Zunino and other participants to seminars, workshops and conferences at EMLyon, Université Côte d'Azur, Waseda University, Warsaw Economic Seminar, 2018 ASFEE Meeting in Nice, 2018 ESA World Meeting in Berlin, 2019 ECBA Workshop in Paris, and 2020 ESA Global Online Meeting for valuable comments. Adam Zylbersztejn acknowledges VISTULA Fellowship.

References

Abeler, J., A. Becker, and A. Falk (2014). Representative evidence on lying costs. Journal of Public Economics 113, 96–104.

Anbarci, N., N. Feltovich, and M. Y. Gürdal (2019). What makes communication informative? Working paper, Monash University.

Balliet, D. (2010). Communication and cooperation in social dilemmas: A meta-analytic review. Journal of Conflict Resolution 54(1), 39–57.

Belot, M., V. Bhaskar, and J. van de Ven (2010). Promises and cooperation: Evidence from a tv game show. Journal of Economic Behavior & Organization 73(3), 396–405.

Belot, M., V. Bhaskar, and J. Van De Ven (2012). Can observers predict trustworthiness? Review of Economics and Statistics 94(1), 246–259.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity, and social history. Games and Economic Behavior 10(1), 122–142.

Bicchieri, C. and A. Lev-On (2007). Computer-mediated communication and cooperation in social dilemmas: an experimental analysis. Politics, Philosophy & Economics 6(2), 139–168.

Bicchieri, C., A. Lev-On, and A. Chavez (2010). The medium or the message? communication relevance and richness in trust games. Synthese 176(1), 125–147.

Bochet, O., T. Page, and L. Putterman (2006). Communication and punishment in voluntary contribution experiments. Journal of Economic Behavior & Organization 60(1), 11–26.

Bochet, O. and L. Putterman (2009). Not just babble: Opening the black box of communication in a voluntary contribution experiment. European Economic Review 53(3), 309–326.

Bolton, G. E., E. Katok, and A. Ockenfels (2004). How effective are electronic reputation mechanisms? an experimental investigation. Management Science 50(11), 1587–1602.

Bonnefon, J.-F., A. Hopfensitz, and W. De Neys (2013). The modular nature of trustworthiness detection. Journal of Experimental Psychology: General 142(1), 143.

Bonnefon, J.-F., A. Hopfensitz, and W. De Neys (2017). Can we detect cooperators by looking at their face? Current Directions in Psychological Science 26(3), 276–281.

Bouas, K. S. and S. S. Komorita (1996). Group discussion and cooperation in social dilemmas. Personality and Social Psychology Bulletin 22(11), 1144–1150.

Bracht, J. and N. Feltovich (2009). Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game. Journal of Public Economics 93(9-10), 1036–1044.

Brosig, J., J. Weimann, and A. Ockenfels (2003). The effect of communication media on cooperation. German Economic Review 4(2), 217–241.

Buchan, N. R., R. T. Croson, and S. Solnick (2008). Trust and gender: An examination of behavior and beliefs in the investment game. Journal of Economic Behavior & Organization 68(3-4), 466–476.

Camerer, C. F. (2003). Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press.

Cameron, A. C. and P. K. Trivedi (2005). Microeconometrics: Methods and Applications. Cambridge university press.

Centorrino, S., E. Djemai, A. Hopfensitz, M. Milinski, and P. Seabright (2015). Honest signaling in trust interactions: Smiles rated as genuine induce trust and signal higher earning opportunities. Evolution and Human Behavior 36(1), 8–16.

Charness, G. (2000). Self-serving cheap talk: A test of aumann's conjecture. Games and Economic Behavior 33(2), 177–194.

Charness, G. and M. Dufwenberg (2006). Promises and partnership. Econometrica 74(6), 1579–1601.

Charness, G. and M. Dufwenberg (2010). Bare promises: An experiment. Economics Letters 107(2), 281–283.

Charness, G. and U. Gneezy (2008). What's in a name? anonymity and social distance in dictator and ultimatum games. Journal of Economic Behavior & Organization 68(1), 29–35.

Chen, J. and D. Houser (2017). Promises and lies: can observers detect deception in written messages. Experimental Economics 20(2), 396–419.

Chovil, N. and A. J. Fridlund (1991). Why emotionality cannot equal sociality: Reply to buck. Journal of Nonverbal Behavior 15(3), 163–167.

Christenfeld, N. (1995). Does it hurt to say um? Journal of Nonverbal Behavior 19(3), 171–186.

Cohn, A., T. Gesche, and M. Maréchal (2018). Honesty in the digital age. Working Paper 280, Department of Economics, University of Zurich.

Conrads, J. and S. Lotz (2015). The effect of communication channels on dishonest behavior. Journal of Behavioral and Experimental Economics 58, 88–93.

Conrads, J. and T. Reggiani (2017). The effect of communication channels on promise-making and promise-keeping: experimental evidence. Journal of Economic Interaction and Coordination 12(3), 595–611.

Cooper, R., D. V. DeJong, R. Forsythe, and T. W. Ross (1992). Communication in coordination games. The Quarterly Journal of Economics 107(2), 739–71.

Crawford, V. (1998). A survey of experiments on communication via cheap talk,. Journal of Economic Theory 78(2), 286–298.

De Neys, W., A. Hopfensitz, and J.-F. Bonnefon (2015). Adolescents gradually improve at detecting trustworthiness from the facial features of unknown adults. Journal of Economic Psychology 47, 17–22.

Duffy, J. and N. Feltovich (2002). Do actions speak louder than words? an experimental comparison of observation and cheap talk. Games and Economic Behavior 39(1), 1–27.

Duffy, J. and N. Feltovich (2006). Words, deeds, and lies: Strategic behaviour in games with multiple signals. Review of Economic Studies 73(3), 669–688.

Eckel, C. C. and R. Petrie (2011). Face value. American Economic Review 101(4), 1497–1513.

Ekman, P. (2006). Darwin and facial expression: A century of research in review. Malor Books, An imprint of The Institute for the Study of Human Knowledge.

Ellingsen, T. and M. Johannesson (2004). Promises, threats and fairness. The Economic Journal 114(495), 397–420.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global evidence on economic preferences. The Quarterly Journal of Economics 133(4), 1645–1692.

Falk, A., A. Becker, T. J. Dohmen, D. Huffman, and U. Sunde (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. Discussion paper 9674, IZA, Institute of Labor Economics.

Farrell, H. and J. Knight (2003). Trust, institutions, and institutional change: Industrial districts and the social capital hypothesis. Politics & Society 31(4), 537–566.

Fetchenhauer, D., T. Groothuis, and J. Pradel (2010). Not only states but traits–humans can identify permanent altruistic dispositions in 20 s. Evolution and Human Behavior 31(2), 80–86.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10(2), 171–178.

Gangadharan, L., N. Nikiforakis, and M. C. Villeval (2017). Normative conflict and the limits of self-governance in heterogeneous populations. European Economic Review 100, 143–156.

Greif, A. (1993). Contract enforceability and economic institutions in early trade: The maghribi traders' coalition. The American Economic Review 83(3), 525–548.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. Journal of the Economic Science Association 1(1), 114–125.

Greiner, B., W. Güth, et al. (2012). Social communication and discrimination: a video experiment. Experimental Economics 15(3), 398–417.

He, S., T. Offerman, and J. van de Ven (2016). The sources of the communication gap. Management Science 63(9), 2832–2846.

Houser, D. and E. Xiao (2011). Classification of natural language messages using a coordination game. Experimental Economics 14(1), 1–14.

Ismayilov, H. and J. Potters (2016). Why do promises affect trustworthiness, or do they? Experimental Economics 19(2), 382–393.

Johnson, N. D. and A. A. Mislin (2011). Trust games: A meta-analysis. Journal of Economic Psychology 32(5), 865–889.

Knack, S. and P. Keefer (1997). Does social capital have an economic payoff? a cross-country investigation. The Quarterly Journal of Economics 112(4), 1251–1288.

Marlow, S. L., C. N. Lacerenza, and E. Salas (2017). Communication in virtual teams: a conceptual framework and research agenda. Human Resource Management Review 27(4), 575 – 589.

Meyerson, D., K. Weick, and R. Kramer (1996). Swift trust and temporary groups. In R. Kramer and T. Tyler (Eds.), Trust in Organizations: Frontiers of Theory and Research. Sage Publications: Thousand Oaks, CA.

Naquin, C. E., T. R. Kurtzberg, and L. Y. Belkin (2010). The finer points of lying online: E-mail versus pen and paper. Journal of Applied Psychology 95(2), 387.

Ostrom, E., J. Walker, and R. Gardner (1992). Covenants with and without a sword: Self-governance is possible. American Political Science Review 86(2), 404–417.

Rideout, V. and M. B. Robb (2018). Social media, social life: Teens reveal their experiences. Research report, Common Sense.

Rockmann, K. W. and G. B. Northcraft (2008). To be or not to be trusted: The influence of media richness on defection and deception. Organizational Behavior and Human Decision Processes 107(2), 106 – 122.

Rousseau, D. M., S. B. Sitkin, R. S. Burt, and C. Camerer (1998). Not so different after all: A cross-discipline view of trust. Academy of management review 23(3), 393–404.

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. Rationality and Society 7(1), 58–92.

Schwartz, S., E. Spires, and R. Young (2019). Why do people keep their promises? a further investigation. Experimental Economics 22(2), 530–551.

Servátka, M., S. Tucker, and R. Vadovič (2011). Words speak louder than money. Journal of Economic Psychology 32(5), 700–709.

Simpson, B., R. Willer, and A. Harrell (2017). The enforcement of moral boundaries promotes cooperation and prosocial behavior in groups. Nature Scientific Reports 7, 42844.

Tognetti, A., V. Durand, M. Barkat-Defradas, and A. Hopfensitz (2020). Does he sound cooperative? acoustic correlates of cooperativeness. British Journal of Psychology forthcoming.

Turmunkh, U., M. J. van den Assem, and D. Van Dolder (2019). Malleable lies: Communication and cooperation in a high stakes tv game show. Management Science 65(10), 4795–4812.

Valley, K. L., J. Moag, and M. H. Bazerman (1998). A matter of trust: Effects of communication on the efficiency and distribution of outcomes. Journal of Economic Behavior & Organization 34(2), 211 – 238.

Van den Assem, M. J., D. Van Dolder, and R. H. Thaler (2012). Split or steal? cooperative behavior when the stakes are large. Management Science 58(1), 2–20.

van Leeuwen, B., C. N. Noussair, T. Offerman, S. Suetens, M. van Veelen, and J. van de Ven (2017). Predictably angry – facial cues provide a credible signal of destructive behavior. Management Science 64, 3352–3364.

Vanberg, C. (2008). Why do people keep their promises? an experimental test of two explanations. Econometrica 76(6), 1467–1480.

Vogt, S., C. Efferson, and E. Fehr (2013). Can we see inside? predicting strategic behavior given limited information. Evolution and Human Behavior 34(4), 258–264.

Woike, J. K. and P. Kanngiesser (2019). Most people keep their word rather than their money. Open Mind 3, 68–88.

Zylbersztejn, A., Z. Babutsidze, and N. Hanaki (2020). Preferences for observable information in a strategic setting: An experiment. Journal of Economic Behavior & Organization 170, 268–285.