



**HAL**  
open science

# Nonverbal content and swift trust: An experiment on digital communication

Zakaria Babutsidze, Nobuyuki Hanaki, Adam Zylbersztejn

► **To cite this version:**

Zakaria Babutsidze, Nobuyuki Hanaki, Adam Zylbersztejn. Nonverbal content and swift trust: An experiment on digital communication. 2020. halshs-02483343

**HAL Id: halshs-02483343**

**<https://shs.hal.science/halshs-02483343>**

Preprint submitted on 18 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

WP 2008 – February 2020

## Nonverbal content and swift trust: An experiment on digital communication

Zakaria Babutsidze, Nobuyuki Hanaki, Adam Zylbersztejn

### Abstract:

We experimentally study the effect of the mode of digital communication on the emergence of swift 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 interaction partners.

### Keywords:

Digital communication, Trust, Hidden action, Nonverbal content, Principal-agent relationship, Promises

### JEL codes:

C72, D83

# Nonverbal content and swift trust: An experiment on digital communication\*

Zakaria Babutsidze<sup>†</sup>

Nobuyuki Hanaki<sup>‡</sup>

Adam Zylbersztein<sup>§</sup>

February 17, 2020

## Abstract

We experimentally study the effect of the mode of digital communication on the emergence of swift 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 interaction partners.

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

**JEL Code:** C72, D83

---

\*This is a revised version of GATE WP 1909 entitled “Digital communication and swift trust”. 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<sup>JEDI</sup>* (ANR-15-IDEX-01) and LABEX CORTEX (ANR-11-LABX-0042) as a part of Université de Lyon (ANR-11-IDEX-007). Ynès Bouamoud, Imen Bouhleb, Maxime David, Carla Lapointe, Ismaël Rafai, Charlotte Saucet, Ann Thampi, Vincent Théroude, Mira Toumi, Dorian Vincileoni, and Charlene Wojerz have provided quality research assistance. Quentin Thévenet has provided valuable assistance with software programming. We thank Giuseppe Attanasi, Brice Corgnet, John Duffy, Martin Dufwenberg, Astrid Hopfensitz, Nicolas Jacquemet, Annett John, Michal Krawczyk, Charles Noussair, 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, and 2019 ECBA workshop in Paris for valuable comments.

<sup>†</sup>SKEMA Business School, Université Côte d’Azur (GREDEG) and OFCE, Sciences Po Paris

<sup>‡</sup>Institute of Social and Economic Research, Osaka University, Japan

<sup>§</sup>Univ Lyon, Université Lyon 2, GATE UMR 5824, F-69130 Ecully, France.

# 1 Introduction

Trust is of utmost value for economic activities. 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).<sup>1</sup> 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.<sup>2</sup>

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 the verbal components (e.g., verbal content and style of the message), but also the nonverbal one (e.g., a speaker’s facial displays, voice, body movements).<sup>3</sup> However, recent

---

<sup>1</sup>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).

<sup>2</sup>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 retransferred in the opposite direction. See Johnson and Mislin (2011) for a meta-analytic study 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.

<sup>3</sup>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 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

IT revolution 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).

Herein, we focus on the role digital communication in building swift trust in human interactions. The main question we are asking in this paper is: does the performance of a digital communication tool in inducing swift trust depend on its capacity to transmit nonverbal content?<sup>4</sup> 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 link between the richness of nonverbal content transmitted through a communication channel and the subsequent emergence of swift trust. We achieve this by inserting the recordings made in prior face-to-face interactions into our three communication treatments.

---

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.

<sup>4</sup>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 broadly, 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). 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.

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 one. 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.<sup>5</sup> We confirm this finding for a large set of communication protocols. Furthermore, we extend this literature by shedding light on the role on nonverbal content of communication, as well as its interplay with the verbal content. Our main findings are as follows. In aggregate, plain text messages entail 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 that these aggregate effects are moderated by the verbal content of communication. Including auditory cues generates a statistically significantly increases trust levels when the message conveys a promise to cooperate. However, in the the absence of such a promise, only the combination of auditive and visual cues significantly increases trust levels compared to the baseline textual communication. Finally, in the additional analyses we also relate our investigation of the role of nonverbal content of communication to the experimental literature on social distance. Following some of the previous experiments on strategic communication, we 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 suffice to identify the other person’s gender). Adding nonverbal

---

<sup>5</sup>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.

content based on auditory cues helps close this gap. However, it also leaves females more trusting 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 non-verbal content of communication on human trust, but also to the heterogeneity of this effect. It may be sensitive not only to the factors that are endogeneous to the process of communication (such as the verbal content that is transmitted by the sender alongside the nonverbal content), but also to the exogeneous ones (such as the players' gender).

## 2 Empirical strategy

The main methodological objective of our study is to provide an experimental design allowing 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 swift 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.<sup>6</sup> However,

---

<sup>6</sup>This echoes the findings from a meta-analytic study Balliet (2010). This study distinguishes between two forms

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.<sup>7</sup> Thus, behaviors observed in these studies may stem not only from an exogenous variation in the 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 cause underlying 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 experimental economics 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

---

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.

<sup>7</sup>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.



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.<sup>8</sup> 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.<sup>9</sup> 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 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

---

<sup>8</sup>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.

<sup>9</sup>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  $p = 0.004$ .

communication on trust.<sup>10</sup> 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 the 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.

### 3 Experimental design and procedures

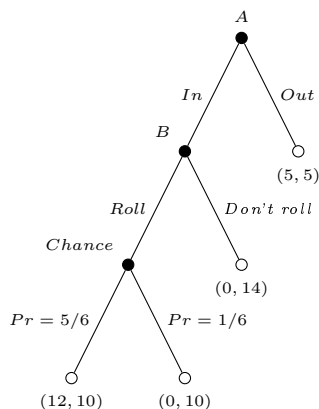
Our three main experimental treatments are based on an earlier set of video recordings from the pre-play face-to-face communication from agents to principals in a classic hidden action game by Charness and Dufwenberg (2006). Participants acting as agents in those earlier sessions with face-to-face communication 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 bare content 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.<sup>11</sup> Below, we first outline the experimental game, then the face-to-face experiment used to gather the recordings, and then our three subsequent experimental treatments that exploit those aforementioned recordings.

---

<sup>10</sup>We also 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.

<sup>11</sup>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



### 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 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 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. 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.<sup>12</sup> Below, we overview the implementation of this hidden action game.

**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 row. 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

---

<sup>12</sup>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.

staff member rolls a die in private and marks the outcome on player B's sealed envelope.

At the end of the experiment, player As and 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.

**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.<sup>13</sup> 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 (including inarticulate sounds) are displayed on player As' computer screens for the amount of time equal to the duration of the corresponding recording. 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 a large university in a medium-sized city in France, between April 2017 and February 2018. All treatments were implemented through a between-subject design. In

---

<sup>13</sup>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.

addition to 83 subjects who participated in F2F experiment, a total of 217 subjects participated in VIDEO (62), AUDIO (70), and TEXT (85) treatments.<sup>14</sup> 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*.<sup>15</sup> There is no significant difference between the duration of the 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 ( $N = 41$ ) which also controls for session fixed effects<sup>16</sup> 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,  $p = 0.001$ ). In the additional analyses reported in Appendix F, we also offer a screening exercise exploiting the

---

<sup>14</sup>Number of participants in VIDEO, AUDIO, and TEXT varies due to the natural variation in show-up rates across sessions.

<sup>15</sup>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 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) and their classification may be found in the Appendix E.

<sup>16</sup>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 ( $p = 0.287$ ).

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.<sup>17</sup>

## 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.<sup>18</sup>

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.<sup>19</sup> Overall, we observe similar trust rates in VIDEO and

---

<sup>17</sup>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 ( $9 \times 3$ ), both *p*-values are computed using Monte Carlo simulations.

<sup>18</sup>As a robustness check, Table 8 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.

<sup>19</sup>See Appendix H for details (as well as caveats) of the comparisons of these three treatments with the F2F

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

	Rate (SE)	$p$
1[ <i>VIDEO</i> ] ( $\alpha_1$ )	0.490 (0.032)	< 0.001
1[ <i>AUDIO</i> ] ( $\alpha_2$ )	0.444 (0.029)	< 0.001
1[ <i>TEXT</i> ] ( $\alpha_3$ )	0.365 (0.026)	< 0.001

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

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

(b) Comparing trust rates from Fig. 2a

	Rate (SE)	$p$
1[ <i>VIDEO, NON - PROMISE</i> ] ( $\beta_1$ )	0.420 (0.040)	< 0.001
1[ <i>AUDIO, NON - PROMISE</i> ] ( $\beta_2$ )	0.342 (0.035)	< 0.001
1[ <i>TEXT, NON - PROMISE</i> ] ( $\beta_3$ )	0.301 (0.030)	< 0.001
1[ <i>VIDEO, PROMISE</i> ] ( $\beta_4$ )	0.551 (0.034)	< 0.001
1[ <i>AUDIO, PROMISE</i> ] ( $\beta_5$ )	0.535 (0.034)	< 0.001
1[ <i>TEXT, PROMISE</i> ] ( $\beta_6$ )	0.434 (0.035)	< 0.001

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

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

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

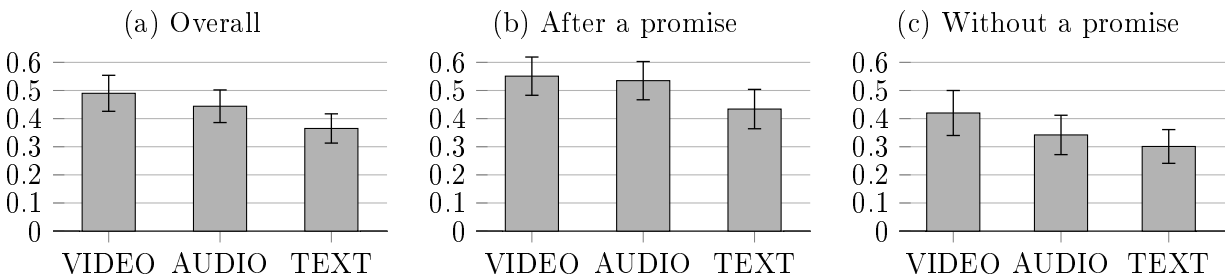
**Note.** (a) and (c): OLS estimates from linear probability models (LPM) regressing the trust indicator variable (= 1 if player A goes *In*, = 0 otherwise) on the all-experiment set of condition indicator variables (= 1 for a given condition, = 0 otherwise). The models are intercept-free, and each coefficient 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 620 observations (62 clusters) in *VIDEO*, 700 (70) in *AUDIO*, and 850 (85) in *TEXT*.



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.

AUDIO (49.0% and 44.4%, respectively;  $p = 0.289$ ) which are, in turn, significantly higher than in TEXT (36.5%, with  $p = 0.003$  vs. VIDEO, and  $p = 0.042$  vs. AUDIO).<sup>20</sup>

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.<sup>21</sup>

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 different to 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 effects of a player B's promise to *Roll* on the overall trust rate are positive (13 percentage points in VIDEO and TEXT, and 19 percentage points in AUDIO)

---

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.

<sup>20</sup>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. 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 ( $p = 0.241$ ).

<sup>21</sup>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, and find no evidence of a correlation in players' decisions.

and statistically significant in all the treatments ( $p < 0.001$ ). Furthermore, these effects do not vary significantly across the three treatments ( $p = 0.402$ ,  $F$ -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? 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.<sup>22</sup> 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.

The main result of this additional analysis 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 ( $p = 0.953$ ).

---

<sup>22</sup>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.

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

	VIDEO		AUDIO		TEXT	
	<i>ME, (SE)</i>	<i>p</i>	<i>ME, (SE)</i>	<i>p</i>	<i>ME, (SE)</i>	<i>p</i>
1[ <i>Promise</i> ]	0.102 (0.030)	0.001	0.171 (0.037)	< 0.001	0.099 (0.031)	0.002
Player A's perception of player B's characteristics (1-8 scale):						
beauty	-0.016 (0.014)	0.232	-0.010 (0.013)	0.449		
intelligence	0.033 (0.018)	0.070	0.019 (0.020)	0.352	0.066 (0.010)	< 0.001
trustworthiness	0.088 (0.013)	< 0.001	0.058 (0.015)	< 0.001	0.033 (0.009)	< 0.001
self-confidence	0.009 (0.012)	0.465	-0.001 (0.012)	0.920	-0.007 (0.009)	0.461
Player A's characteristics:						
SVO angle	-0.002 (0.002)	0.249	< 0.001 (0.002)	0.928	< 0.001 (0.002)	0.954
1[ <i>CRT</i> > 0]	0.020 (0.066)	0.759	0.069 (0.058)	0.233	-0.052 (0.067)	0.437
RMET score	0.003 (0.007)	0.680	-0.004 (0.005)	0.469	0.002 (0.006)	0.782
GP investment	0.001 (0.001)	0.285	0.003 (0.001)	0.013	< 0.001 (0.001)	0.988
SOEP trust 1	-0.074 (0.053)	0.161	-0.072 (0.053)	0.172	-0.068 (0.054)	0.207
SOEP trust 2	-0.003 (0.050)	0.953	0.097 (0.035)	0.006	0.084 (0.046)	0.072
SOEP trust 3	0.081 (0.072)	0.260	-0.016 (0.068)	0.813	0.116 (0.074)	0.118
Political views	-0.001 (0.017)	0.943	0.003 (0.013)	0.840	0.003 (0.010)	0.767
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.

### 4.3 Additional analyses: social distance and gender effects in trust

Following Bohnet et al. (1999), Hoffman et al. (1996), and He et al. (2016), we define social distance as the extent to which the other person can be identified by the decision-maker.<sup>23</sup> Our experimental conditions vary social distance in a meaningful way: TEXT provides the lowest degree of identification of a player B (restricted to the content of the statement), AUDIO improves it thanks to the additional auditive cues, and VIDEO further enhances it by providing both visual and auditive cues about a player B. In particular, one crucial difference between TEXT and the remaining treatments is that in the former, a player B’s gender remains mostly unknown to player As.<sup>24</sup> In this part of our analysis, we consider the availability of information about the counterpart’s gender as a factor reducing social distance in the interaction, and explore how it affects trust in different communication environments.

We note that the existing experimental literature on communication lacks a consensus on the relationship between (the availability of cues about the other person’s) 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 ( $p = 0.479$ ).<sup>25</sup> Thus, we interpret any

---

<sup>23</sup>We also note that some other experimental studies cast doubt on whether the degree of social distance may behavior in economic games (Dufwenberg and Muren, 2006; Deck et al., 2013).

<sup>24</sup>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*.

<sup>25</sup>This result stems from a logit regression with session fixed effects, and holds (marginal effect of -0.088,  $p = 0.534$ )

Table 3: Trust and gender: aggregate results

<i>Ratio trust</i>	VIDEO		AUDIO		TEXT	
	<i>F</i>	<i>M</i>	<i>F</i>	<i>M</i>	<i>F</i>	<i>M</i>
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 rate at which that player trusted male (female) player Bs. The average values in the sample are conditioned treatment and player A's gender.

difference in player As' behavior that stems from player Bs' gender as a behavioral bias.

We further investigate if there is a gender difference in player As' trust, and whether the trust in our experiment is gender-biased. To answer these questions, let us 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 for female).<sup>26</sup> The sample means of those individual ratios are summarized in Table 3.

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

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

where  $1[Female\_A_i] = 1$  if an *i*th player A is a female, and = 0 otherwise. Since both *Ratio\\_trust<sub>iF</sub>* and *Ratio\\_trust<sub>iM</sub>* 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.

The main results are as follows. In TEXT (where player B's gender is predominantly unob-  


---

 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,  $p = 0.001$ ).

<sup>26</sup>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 14 for these data gives  $p = 0.287$ .

Table 4: Trust and gender: seemingly unrelated regressions

<i>Ratio trust</i>	VIDEO		AUDIO		TEXT	
	<i>F</i>	<i>M</i>	<i>F</i>	<i>M</i>	<i>F</i>	<i>M</i>
Intercept ( $\gamma_0$ )	0.541 <sup>a</sup> (0.065)	0.496 <sup>a</sup> (0.061)	0.535 <sup>a</sup> (0.067)	0.464 <sup>a</sup> (0.056)	0.430 <sup>a</sup> (0.046)	0.437 <sup>a</sup> (0.045)
1[ <i>Female_A</i> ] ( $\gamma_1$ )	-0.052 (0.080)	-0.034 (0.076)	-0.031 (0.081)	-0.087 (0.069)	-0.123 <sup>b</sup> (0.060)	-0.113 <sup>c</sup> (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.

servable 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  $\gamma_0$  as well as  $\gamma_0 + \gamma_1$  are not different in equations *Ratio trust F* and *Ratio trust M* ( $p = 0.871$  for  $\gamma_0$  and  $p = 0.629$  for  $\gamma_0 + \gamma_1$ ). In AUDIO, male player As' trust does not vary as a function of player B's gender (coefficients  $\gamma_0$  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 swift trust? Overall, our experimental data highlight an interplay between verbal and nonverbal content of communication in the process of inducing swift 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 swift trust in human interactions. In aggregate terms, providing auditory cues about the speaker seems to play a key role in inducing swift 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.

In addition, we discuss whether and how digital communication protocols may give rise to gender differences in trust, depending on the amount of transmitted nonverbal cues. We relate these findings to the concept of social distance (Bohnet et al., 1999; Hoffman et al., 1996) which

has been previously studied in the context of (face-to-face) communication by He et al. (2016) and which can be defined as the extent to which the other person can be identified by the decision-maker. 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.<sup>27</sup> Providing auditory cues that reveal the sender’s gender, but otherwise provide only a limited identification of that person, 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, ensuring a high degree of identification of the other person.

In the experiment by He et al. (2016), subjects keep a direct visual contact (except for the control condition), while the scope of communication varies across treatments: from no communication at all to restricted communication (no promise-making) to free communication (including promise-making). Their data suggest that social distance does not play an important role in explaining the effect of communication on cooperation in a social dilemma; rather, that effect is due commitment building and type recognition (which is particularly pronounced under face-to-face communication). Our experimental design and the conclusions thereof are somewhat different. We study a free communication environment with a highly homogeneous message content. We observe that the extent to which the mode of digital communication allows the decision-maker to identify the other person plays a role in the emergence of trust towards that person. If the available cues are too scarce to fully identify the other person, female decision-makers may be either relatively wary of trusting others, or focus on (and seek commonalities in) a characteristic that may be correctly identified (i.e. the other person’s gender).

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

---

<sup>27</sup>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.



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.

## References

- Abeler, J., A. Becker, and A. Falk (2014). Representative evidence on lying costs. Journal of Public Economics 113, 96–104.
- Ashraf, N., I. Bohnet, and N. Piankov (2006). Decomposing trust and trustworthiness. Experimental Economics 9(3), 193–208.
- Balliet, D. (2010). Communication and cooperation in social dilemmas: A meta-analytic review. Journal of Conflict Resolution 54(1), 39–57.
- Baron-Cohen, S., D. C. Bowen, R. J. Holt, C. Allison, B. Auyeung, M. V. Lombardo, P. Smith, and M.-C. Lai (2015). The “reading the mind in the eyes” test: Complete absence of typical sex difference in ~400 men and women with autism. PLoS ONE 10(8), e0136521.
- Baron-Cohen, S., S. Wheelwright, J. Hill, Y. Raste, and I. Plumb (2001). The “reading the mind in the eyes” test revised version: a study with normal adults, and adults with asperger syndrome or high-functioning autism. The Journal of Child Psychology and Psychiatry and Allied Disciplines 42(2), 241–251.
- 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.
- Bohnet, I., B. S. Frey, et al. (1999). The sound of silence in prisoner’s dilemma and dictator games. Journal of Economic Behavior & Organization 38(1), 43–57.
- 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.
- Brañas-Garza, P., T. García-Muñoz, and R. Hernán González (2012). Cognitive effort in the beauty contest game. Journal of Economic Behavior & Organization 83(2), 254–260.
- 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.

- 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.
- Corgnet, B., A. M. Espín, and R. Hernán-González (2015). The cognitive basis of social behavior: cognitive reflection overrides antisocial but not always prosocial motives. Frontiers in Behavioral Neuroscience 9, 287.
- Cox, J. C. (2004). How to identify trust and reciprocity. Games and Economic Behavior 46(2), 260–281.
- 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.
- Deck, C., M. Servátka, and S. Tucker (2013). An examination of the effect of messages on cooperation under double-blind and single-blind payoff procedures. Experimental Economics 16(4), 597–607.
- 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.
- Dufwenberg, M. and A. Muren (2006). Gender composition in teams. Journal of Economic Behavior & Organization 61(1), 50–54.
- 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. Grootuis, 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.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. Biometrika 80(1), 27–38.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10(2), 171–178.
- Frederick, S. (2005). Cognitive reflection and decision making. Journal of Economic Perspectives 19(4), 25–42.

- 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.
- Gneezy, U. and J. Potters (1997). An experiment on risk taking and evaluation periods. The Quarterly Journal of Economics 112(2), 631–645.
- 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.
- Hanaki, N., N. Jacquemet, S. Luchini, and A. Zylbersztein (2016). Fluid intelligence and cognitive reflection in a strategic environment: Evidence from dominance-solvable games. Frontiers in Psychology 7, 1188.
- He, S., T. Offerman, and J. van de Ven (2016). The sources of the communication gap. Management Science 63(9), 2832–2846.
- Heinze, G. and M. Schemper (2002). A solution to the problem of separation in logistic regression. Statistics in Medicine 21(16), 2409–2419.
- Hoffman, E., K. McCabe, and V. L. Smith (1996). Social distance and other-regarding behavior in dictator games. The American Economic Review 86(3), 653–660.
- 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.
- Murphy, R. O., K. A. Ackermann, and M. J. Handgraaf (2011). Measuring social value orientation. Judgment and Decision Making 6, 771–781.
- 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.
- 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. Spire, 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.
- Tavakol, M. and R. Dennick (2011). Making sense of cronbach's alpha. International Journal of Medical Education 2, 53.
- Tognetti, A., V. Durand, M. Barkat-Defradas, and A. Hopfensitz (2020). Does he sound cooperative? acoustic correlates of cooperativeness. British Journal of Psychology forthcoming.
- 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.
- Veazie, P. J. (2006). When to combine hypotheses and adjust for multiple tests. Health Services Research 41, 804–818.
- 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.
- Wolke, 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.



# Appendix

## A Experimental instructions

### A.1 Preliminary instructions given to all subjects in all the treatments

You are about to take part in an experiment in which you can earn money. The amount of your gains will depend on your decisions, as well as on the decisions made by other participants. In addition, you will receive a fixed fee of [*5 for player As, 10 for player Bs*] EUR for completing the experiment. Your total earnings will be paid privately in cash at the end of the experiment.

The experiment consists of several parts. Each part will involve tasks the rules of which will be explained to you in due time. It is crucial that you understand and obey the rules of this experiment. Violation of these rules might result in an exclusion from the experiment and all payments. Please raise your hand whenever you have questions or need assistance.

**All the information you provide, as well as the amount of your gains from this experiment, will remain strictly confidential and anonymous.**

We would now like to ask you to answer a series of preliminary questions. You will answer these questions using the interface on your computer screen. Some of these questions will generate monetary gains. These gains will be determined and added to your overall earnings at the end of the experiment.

*Note: the following instructions were given only to player Bs in F2F for the preliminary recordings, not reported in the paper.*

Now, we would like to take a picture and video recording of you.

First, you will be asked to stand by the wall and look into the camera. Please, try to keep a neutral facial expression.

Second, you will be asked to read aloud the content display on the screen in front of you. While reading, you will be video recorded.

All pictures and video recordings produced during this experiment will only serve strictly scientific purposes of this research project. They may be used in other experimental sessions related to this research project.

## **A.2 Specific instructions for F2F treatment**

### **Rules of the game**

You will now play a game with monetary stakes. The rules of the game are as follows.

The game is played by two players: player A and player B. Each player must choose between two possible actions. Player A chooses between actions “Lef” and “Righ”. Player B chooses whether she want a six-sided die to be rolled (action “Roll”) or not (action “Don’t roll”).

**You will play the role of player [A for player As, B for player Bs]**

Each players’ payoff depends on the actions chosen by herself as well as the other player:

- if player A chooses “Lef”, then regardless of player Bs’ choice:
  - player A’s payoff is 5 EUR and player B’s payoff is 5 EUR;
- if player A chooses “Righ” and player B chooses “Don’t roll”:

- player A’s payoff is 0 EUR and player B’s payoff is 14 EUR;
- if player A chooses “Right” and player B chooses “Roll”:
  - if the number of on the die is between 1 and 5, then player A’s payoff is 12 EUR and player B’s payoff is 10 EUR;
  - if if the number of on the die is 6 player A’s payoff is 0 EUR and player B’s payoff is 10 EUR;

### How the game proceeds

The game will consist of six identical rounds.

At the beginning of a round, one player B is asked to enter the room in which there are six players As. Player As are separated one from another and are not allowed to talk.

Player B is then placed in front of player As and remains silent. Then, player B is allowed to talk for no longer than 20 seconds, and then asked to leave player As’ room. While talking, player B is video recorded and should look straight into the camera.

Once player B leaves player As’ room:

- player B makes a decision in a separate room. Player B privately and individually indicates her decision (either “Roll” or “Don’t roll”) on a separate answer sheet, puts it in an envelope and seals the envelope. The experimenter collects the envelope and player B leaves the room. Then, the experimenter privately rolls a six-sided die and marks the result on the envelope (without opening it). The outcome of the die roll will only be taken into account if player A’s decision is “Left” and player B’s decision is *Roll*.
- each player A privately and individually indicates her decision (either “Left” or “Right”) on a separate answer sheet, puts it in an envelope and seals the envelope. Then, all the envelopes are collected by the experimenter. Player As are either asked to remain silent and await the next player B, or informed that the experiment is over and given further instructions about their payment.

No envelope will be opened before the end of the experiment.

At the end of the experiment, each player A is anonymously and randomly matched with one player B. The outcome of the game for each pair of players is determined by the decisions made by both players (and also by the outcome of the die roll if the decisions in a pair are “Right’ and “Roll’) in the round in which the player B was in player As’ room. Players are only informed about their personal payoffs, and not about the payoffs of or the decisions made by other players, or about the outcome of the die roll.

### **Additional information**

Note that this set of instructions is provided to and read by each player A and each player B. Furthermore, player Bs cannot communicate between themselves at any point of the experiment. The same applies to the communication between player As.

**You will play the role of player [A for player As, B for player Bs]**

### **A.3 Specific instructions for VIDEO, AUDIO, and TEXT**

*Note: Below, the parts of instructions that are distinct for each treatment are marked with “(treatment’s name)”. Other parts are common to all three treatments.*

(VIDEO) In this part of the experiment, you will watch a series of video recordings. In each recording, you will see a person speaking to other people. You will also answer a series of questions after each recording; some of them will determine your gain from this experience.

(AUDIO) In this part of the experiment, you will listen to a series of audio recordings. In each recording, you will hear a person speaking to other people. You will also answer a series of questions after each recording; some of them will determine your gain from this experience.

(TEXT) In this part of the experiment, you will read a series of transcripts. Each transcript comes from a person speaking to other people. You will also answer a series of questions after each recording; some of them will determine your gain from this experience.

**(VIDEO) Video recordings; (AUDIO) Audio recordings; (TEXT) Transcripts**

(VIDEO) Here is some information about the recordings you will be watching. The recording have been made during several laboratory experiment sessions.

(AUDIO) Here is some information about the recordings you will be listening to. The recording have been made during several laboratory experiment sessions.

(TEXT) Here is some information about the transcripts you will read. The transcripts have been made during several laboratory experiment sessions.

In each session, two groups of participants (six players A and six players B) were installed in two different rooms. Participants in each room could not communicate with each other. Players were informed that their decisions and earnings would remain private and anonymous, and would never be disclosed to other participants.

Each session was organized as follows:

1. One by one, player Bs entered the room in which players A were sitting. Then, each player B made a short speech in front of player As.

(VIDEO) All the speeches have been recorded, and you will be watching some of them.

(AUDIO) All the speeches have been recorded, and you will be listening to some of them.

(TEXT) All the speeches have been recorded, and you will read some of their transcripts.

2. After his speech, player B left player As' room, and entered an empty room.
3. After Player B's departure, each player A made a decision ("Left" or "Right") in private and individually. At the same time, player B made a decision ("Roll" or "Don't roll" a die) in privately and individually.
4. Thereafter, player B left the room and waited outside the laboratory until the end of the experiment. Meanwhile, a new player B was entering the players' room A to make a speech. The experiment ended when all the players had completed their task.

At the end of the experiment, each player A was anonymously and randomly matched with a player B. The outcome of the game for each pair of players was determined by the decisions made by both players following player B's speech:

- if player A chose "Left", then regardless of player B's choice:
  - player A's payoff was 5 EUR and player B's payoff was 5 EUR;
- if player A chose "Right" and player B chose "Don't roll":
  - player A's payoff was 0 EUR and player B's payoff was 14 EUR;
- if player A chose "Right" and player B chose "Roll":
  - if the number of on the die was between 1 and 5, then player A's payoff was 12 EUR and player B's payoff was 10 EUR;
  - if the number of on the die was 6, then player A's payoff was 0 EUR and player B's payoff was 10 EUR;

## Your role

(VIDEO) You will now watch ten recordings randomly selected from all the recordings described above. After each recording, you will make a decision: either "Left" or "Right". This decision is important for your final earnings. You will also answer some questions about the person seen in the recording. Finally, either you will proceed to the next recording, or you will be informed that the experiment is over.

At the end of the experiment, one of the recordings you have viewed will be drawn at random. Your earnings will depend on the decision ("Left" or "Right") you have made after watching that recording. It will also depend on the decision previously made by the person from the recording following his or her speech in the laboratory session during which the recording was made:

(AUDIO) You will now hear ten recordings randomly selected from all the recordings described above. After each recording, you will make a decision: either "Left" or "Right". This decision is

important for your final earnings. You will also answer some questions about the person heard in the recording. Finally, either you will proceed to the next recording, or you will be informed that the experiment is over.

At the end of the experiment, one of the recordings you have listened to will be drawn at random. Your earnings will depend on the decision (“Left” or “Right”) you have made after listening to that recording. It will also depend on the decision previously made by the person from the recording following his or her speech in the laboratory session during which the recording was made:

(TEXT) You will now see ten transcripts randomly selected from all the transcripts described above. After each transcript, you will make a decision: either “Left” or “Right”. This decision is important for your final earnings. You will also answer some questions about the person that made the speech presented in the transcript. Finally, either you will proceed to the next transcript, or you will be informed that the experiment is over.

At the end of the experiment, one of the transcripts you have read will be drawn at random. Your earnings will depend on the decision (“Left” or “Right”) you have made after reading that transcript. It will also depend on the decision previously made by the person whose speech was presented in the transcript, during the laboratory session in which that speech was made:

- If you chose “Left”, then whatever the choice of the person from the recording: you earn 5 EUR ;
- If you chose “Right” and the person from the registration has chosen “Don’t roll”: your earn 0 EUR ;
- If you chose “Right” and the person from the recording has chosen “Roll”:
  - if the number on the die was between 1 and 5, then you earn 12 EUR;
  - if the number on the die was 6, then you earn 0 EUR;

## B Additional measures

### B.1 Measures of individual characteristics

The preliminary stage of the experiment consists of basic socio-demographic questions (age, gender, education, major, current occupation, score at the *baccalauréat* exam at the end of high school), as well as a battery of incentivized and non-incentivized computerized tasks designed to measure specific individual characteristics. These tasks are administered in five blocks presented to each individual in random order. In the incentivized tasks, all payoffs were expressed in Experimental Currency Units (ECU), with the exchange rate of  $100 \text{ ECU} = 2.50 \text{ EUR}$ . No feedback on any decision or answer is provided at any stage. Below, we describe the content of those parts (labelled A-E).<sup>28</sup>

**Part A.** Other-regarding preferences are measured using the Social Value Orientation (SVO) slider task by Murphy et al. (2011). In each of six distributional tasks, the decision-maker chooses an allocation of money for himself and for another person among nine possible allocations. These choices are then transformed into a score, labelled as “SVO angle”: the higher the angle, the stronger one’s concerns about others’ welfare. We use the original set of distributional tasks (all the amounts are expressed in ECU), and the strategy method to elicit responses in the role of the decision-maker from each participant. We also inform them that they will be randomly and anonymously matched in pairs at the end of the experiment, that in each pair one person will be randomly chosen as the decision-maker, and that both players’ payoffs from this task will correspond to the decision-maker’s choice in a randomly selected task.

**Part B.** Cognitive skills are measured through standard (3-item) Cognitive Reflection Test (Frederick, 2005) which “*measures cognitive reflectiveness or impulsiveness, respondents’ automatic response versus more elaborate and deliberative thought*” (Brañas-Garza et al., 2012, p.255). The questions are as follows:

---

<sup>28</sup>Table 5 in Appendix D offers a summary of the elicited characteristics. We also provide evidence that the sample has been properly randomized with respect to all of these characteristics across treatments (VIDEO, AUDIO, or TEXT) and roles (player A or player B). Finally, we replicate several results that have been previously documents by the experimental studies employing certain measures used herein. In addition, Table 6 summarizes physical characteristics of the sample of player Bs.



1. *A notebook and a pencil cost 1.10 Euros in total. The notebook costs 1 Euro more than the pencil. How much does the pencil cost?*
2. *If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?*
3. *In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?*

Subjects are informed that this set of three questions should be answered within 30 seconds (although we allow them to provide answers even after this time has elapsed). In this way, subjects can be classified according to their overall score (that is, the total number of correct answers) which can range from 0 to 3. No monetary incentive was provided for giving correct answers to these questions.

**Part C.** The Reading the Mind in the Eyes Test is an advanced test of theory of mind (ToM) developed by Baron-Cohen et al. (2001). As noted by Baron-Cohen et al. (2015), *ToM entails attribution and recognition of mental states in oneself or others [...] and to use such information to make inferences and predict behaviour* (p. 2). The test is non-incentivized and consists of 36 black-and-white photos of people. Each photo is cropped and rescaled so that only the area around the eyes can be seen. Photos are displayed one-by one and the participants are asked to choose the word that best describes what the person in the photo is thinking or feeling among four possible words shown underneath (only one of which is correct). Time is limited to 10 minutes. Alongside a set of precise instructions about the content and duration of the task, the participants go through one trial round with an immediate feedback on the accuracy of their answers; other than that, no feedback is provided.

**Part D.** Risk preferences are captured by the amount invested in the incentivized task based on Gneezy and Potters (1997). A decision-maker holds an initial endowed of 100 ECU, some (or all) of which he can invest in the following lottery: 50% chance of multiplying the investment by the factor of 2.5 times, and 50% chance of losing the invested amount. Any decision-maker who does

not invest the whole amount is considered as being risk averse, and the invested amount is used as a measure of risk aversion.

**Part E.** The non-incentivized measures of individual trust attitudes are based on the German Socio-Economic Panel Study (SOEP), an ongoing longitudinal survey of German households. In what follows, we present them in order of appearance. Variable *SOEP trust 1* is the answer to “In general, one can trust other people” (4 point scale with 1 – “strong disagreement”, and 4 – “strong agreement”); variable *SOEP trust 2* is the answer to “How much do you trust people you just met?” (4 point scale with 1 – “no trust at all”, and 4 – “full trust”); variable *SOEP trust 3* is the average score on 7 items measuring general trust in socio-economic institutions (“How much trust do you have in market economy/large companies/courts/government/parliament/other citizens/banks ”, same scale as for *SOEP trust 2*).<sup>29</sup> Finally, political views are elicited using another SOEP question: “In politics, people often talk about “left” and “right” when describing different political views. When you think about your own political view, how would you rate them on the scale below?” (11 point scale with 0 – “left”, and 10 – “right”).

## B.2 Post-decision questionnaire

After every decision (and before moving on to the next round), player As fill in a short questionnaire (based on Centorrino et al., 2015). In all treatments but TEXT, player As are asked if they know the person with whom they have just interacted. Then, they evaluate that person’s: beauty (inverse 1-8 scale: 1 – “beautiful”, 8 – “not beautiful”), trustworthiness (inverse 1-8 scale: 1 – “trustworthy”, 8 – “untrustworthy”), intelligence (1-8 scale: 1 – “not too intelligent”, 8 – “very intelligent”), and self-confidence (1-8 scale: 1 – “shy”, 8 – “self-confident”). For the sake of the simplicity of exposition, the statistical analysis reverses those scales for beauty and trustworthiness. In F2F, the questionnaire is pen-and-paper. Each completed questionnaire sheet is put in an envelope which is then sealed and collected by a laboratory staff member. In VIDEO, AUDIO, and TEXT the questionnaire is embedded in the computer interface used in those treatments.

---

<sup>29</sup>Cronbach’s  $\alpha = 0.7$  suggests that the individual answers are internally consistent. Tavakol and Dennick (2011) describe internal consistency as the extent to which all the items in a test measure the same concept (here, trust towards others). It is connected to the inter-relatedness of the items within the test (p. 53). In the literature,  $\alpha = 0.7$  is usually considered as acceptable (p. 54).

## C Implementation details

**Recruitment of participants.** We made sure that player Bs were aware of (and give their consent to) the fact that they would be recorded, as well as the future use of those recordings for scientific purposes. The invitation email sent to player Bs included the following information: *During the experiment, your photos and video recordings will be taken. All pictures and video recordings produced during this experiment will only serve strictly scientific purposes of this research project, but they will be used in other experimental sessions related to the research project. Therefore, please register only if you agree that your photos and video recording will taken for this scientific purpose.*<sup>30</sup> Similar information is repeated in the preliminary part of the experimental instructions (see Appendix A). For obvious reasons, the invitation email emphasizes that subjects who want to participate must be fluent French speakers.

In order to minimize the chance of subjects in the opposite roles knowing each other, we tried to separate our subject pools and recruited participants from two different majors (player As: BA students in economics and some other majors with the exception of public administration; player Bs: BA students in public administration, MA students in economics, if necessary completed by students from other departments). Acquaintance between player As and player Bs occurs at a relatively low rate. In F2F, in 14.6% (36 in 246) of times a player A admits to know a player B. In VIDEO and AUDIO, this rate is even lower (1.1% and 1.3%, respectively). We suspect that the difference between F2F and the two other conditions stems from the fact that in F2F subjects in both roles physically participated in the same session. Thus, despite our attempts to limit the correlation between player As and player Bs, factors such as shared courses, overlapping time schedules, or simple word of mouth, may be still be at play. This problem happens to be marginal in VIDEO and AUDIO in which we only recruit player As, and becomes irrelevant in TEXT.

**Gender composition of the experimental sample.** In order to homogenize the audience faced by player Bs in F2F and enhance our control over the experimental environment, we attempted to have gender-balanced groups of player As in each session. We achieved this in 5 sessions out of 7. In session 2, there are 2 males and 4 females, while the opposite holds in session 3. Despite these

---

<sup>30</sup>The recruitment of the participants has been carried out using ORSEE (Greiner, 2015).

small variations, altogether the sample of player As in F2F is gender-balanced. In the remaining treatments, the gender-balance constraint is relaxed. Female player As account for 64.5% of the sample in VIDEO, 67.1% in AUDIO, and 57.1% in TEXT. Overall, this procedure does not produce a statistically significant variations in the gender composition of player As' samples. Comparing the four samples using Fisher's exact test yields  $p = 0.273$ .

**Logistics of the F2F condition.** For the sake of logistics and efficient time management, player Bs arrive 30 minutes prior to player As. First, they are asked to take up several computerized tasks that measure their preferences and characteristics (see Section B.1).<sup>31</sup> Then, they are all led to a waiting room. To avoid any communication or subjects overhearing what others are saying or doing, each participant is seated in a separate cubicle, puts on a headphone and listens to a classical music until further notice. Then, they are taken one by one to a separate room for a mugshot picture and a short, standardized video recording.<sup>32</sup>

Then, each subject is seated back in his cubicle with headphones on. He now listens to an audio file containing the experimental instructions (paper version is also provided). There is a brief comprehension quiz assisted by a laboratory staff member. Finally, he receives additional paper instructions about the upcoming statement in front of player As, as well as a pen and an empty sheet of paper, and is given approximately two minutes to prepare his message.<sup>33</sup> After that, a player B is invited to player As' room where he delivers a statement, leaves for another room, and the game proceeds as explain above in Section 3. The average duration of a message is 26.39 seconds (SD 2.09).

Upon their arrival to the laboratory, player As also take up the set of preliminary questionnaires. Then, they receive and read paper instructions for the experimental game, and finally they fill in a short comprehension quiz. A laboratory staff member then reads aloud all the questions from the quiz along with the corrects answers, and answers any remaining questions. Finally,

---

<sup>31</sup>The computerized parts of the experiment are all programmed in z-tree (Fischbacher, 2007).

<sup>32</sup>Like in van Leeuwen et al. (2017), subjects are asked to read neutral content (a short extract from a printer instruction manual) and keep a neutral face expression. The recording takes about 30 seconds. This information is not part of the present investigation and is not reported in the paper.

<sup>33</sup>Those additional instructions remind the subject about his role in the game; emphasize the fact that the message may affect player A's decisions and, consequently, the subject's gain from the experiment; instruct the subject to avoid making a visual or verbal contact with the experimenter, to aim at communicating with all player As, and not to introduce oneself or give any details about one's own identity.

player As are asked to wait for the arrival of the first player B.

**Recordings.** Player Bs' statements are recorded using a small, non-intrusive video camera set up in the middle of player As' row, right in front of player Bs' zone, so that the perspective in the video camera recording resembles the one of a player A. The camera is always adjusted to the height of player B (so as to capture head, shoulders, and thorax), and to the luminosity in the room. The sake of the quality of the video recordings, the background in player Bs' zone is covered with light canvas. While making a statement, each player B also has a portable microphone attached below their face. The distance between player As and a player B is set to 2.50 meters. The information recorded in the F2F treatment (i.e., player Bs' messages and the content of their decisions sheets and die rolls) is re-used in the remaining experimental conditions: VIDEO, AUDIO, and TEXT.

**Duration and payoffs.** For player As, the experiment takes roughly 1h. At the end of the experiment, subjects are paid according to their earnings in the incentivized pre-experimental tasks, earnings in the experimental game, in addition to the show-up fee of 5 euros. All payments are made privately and in cash. For player Bs, the experiments takes 30 minutes longer, and the show-up fee is increased to 10 euros.

## D Individual characteristics: randomization and key empirical patterns in the sample

Table 5 provides a summary of the individual measures outlined in Section B.1. To check if our five experimental subsamples (player As in F2F, VIDEO, AUDIO, and TEXT, as well as player Bs in VIDEO) have been properly randomized with respect to all of these characteristics, we adopt a parametric procedure that is parsimonious (it involves a single test) and provides a simple way to deal with the risk of false rejection due to multiple comparisons (see Veazie, 2006, for an insightful discussion). We regress each outcome of interest (i.e, SVO angle, CRT and RMET scores, investment in the Gneezy-Potters task, SOEP trust 1, 2, and 3, and political views) on the set of four experimental sample dummies (each set to 1 if an observation comes from a given sample, and 0 otherwise; the fifth sample is the reference). These eight regression models are estimated

Table 5: Individual characteristics measured in the experimental sample: average outcomes

	$N$	SVO angle	CRT score	RMET score	GP inves't	SOEP trust			Polit. views
						1	2	3	
Player As:									
F2F	42	21.613	0.405	24.000	48.643	2.310	1.905	2.255	4.452
VIDEO	62	26.973	0.403	26.790	44.065	2.339	1.984	2.306	4.645
AUDIO	70	28.078	0.529	26.871	43.271	2.443	2.129	2.422	4.143
TEXT	85	26.487	0.353	26.435	48.624	2.412	2.047	2.380	4.094
All player As	259	26.243	0.424	26.243	46.089	2.386	2.031	2.354	4.297
Player Bs:	41	21.797	0.415	25.951	46.512	2.390	2.122	2.303	4.683
All participants:	300	25.635	0.420	26.203	46.147	2.387	2.043	2.347	4.350

as seemingly unrelated regressions (SUR). This, in turn, enables us to perform  $F$ -test of joint insignificance of all between-sample differences (i.e., joint insignificance of all sample dummies) simultaneously across all the SUR models. The resulting joint test of 32 null hypotheses yields  $p = 0.148$ .

We also replicate several results previously documents by the experimental studies employing CRT and RMET. First, we find gender differences in both measurements: males score significantly higher than females in the CRT (0.543 vs. 0.316; two-sided  $t$ -test:  $p = 0.005$ ), and significantly lower in the RMET (25.442 vs. 26.637; two-sided  $t$ -test:  $p = 0.019$ ). This stands in line with the findings by Brañas-Garza et al. (2012), Corgnet et al. (2015), and Hanaki et al. (2016) for the CRT, and by Baron-Cohen et al. (2015) for RMET.

Second, we observe that most (74%) of our 300 participants fail to provide at least one correct answer in the standard CRT (14.4% provide exactly one , 8.5% – two, and 3.1% – three correct answers). Such finding is not new in the literature, and echoes a strongly left-skewed distribution of correct answers in the “lower bound” sample from the seminal study by Frederick (2005).<sup>34</sup>

<sup>34</sup>In Frederick (2005), the respective frequencies are 64%, 21%, 10%, 5% ( $N=138$ ). Similar findings are also reported by Brañas-Garza et al. (2012) – 67%, 23%, 9%, and 1% ( $N = 191$ ), and by Hanaki et al. (2016) – 70%, 16%, 8%, and 6% ( $N = 200$ ).

Table 6: Player Bs' physical characteristics

Mean age	22.51
Male (%)	51.22
Ethnicity (%):	
African	7.32
Caucasian	63.41
Middle-eastern	26.83
South Asian	2.44
Native speaker (%)	80.49
Wearing glasses (%)	29.27
Bearded male (%)	31.71
Female wearing hijab (%)	7.32
Female with clivage exposed (%)	34.15

## E Translated transcripts and classification of player Bs' statements

**Note.** For player As, the experimental instructions use a neutral frame: "Right" or "Left", which is analogous to "In" or "Out" in the game from Figure 1.

1. So, I don't know if you know about the experience that's coming, but you have two choices to make. So, you can choose between "Right" and "Left". So, "Left" is the "safe" choice, that is to say that when you earn 5€, I earn 5€. And "Right" is the risky choice, i.e. you may win 0€ and I may win 14€. On the otherhand, it is also a choice that could make us all earn money, in the sense that you would earn 12€ if I remember correctly, and me 10. So what does that mean? This means you have to have confidence. Well, that's complicated, I understand that, especially since we don't know each other at all, and from what I've read, it's going to be intermingled over there. All I can tell you is that no matter what choice you make, I will choose to roll the die, because in any case this solution will best maximize the profit of all of us. Whether you choose the "safe" solution, that is "Left", or the other

solution - “Right”, anyways, it’s in my best interest to play the die, because no matter what you choose, I will earn money. As for you, you’re actually facing the risk of earning nothing at all. Do what you want, in any case I, do your statistics with those who will come next, I in any case it will go full, full dice. That’s it. *[Classification: promise]*

2. Good morning, everyone. Just to let you know that I will always roll the die in every period. *[Classification: promise]*

3. Dear player As, I encourage you to choose the option “Right”. I commit myself to choose the action “Roll” in order to maximize all earnings, in a fair way, between you and me, since this option allows you to win the maximum in this game, that is 12€. *[Classification: promise]*

4. So you have all read the rules of the game, just like me. I suppose you won... you came here, I’m sorry, to make the most money... you didn’t come here for nothing. So I advise all of you to choose the option “Right”, because I will choose the option... "win"... no, the option “Roll”, sorry. That’s what it takes to make the most money. OK, I’m done. *[Classification: promise]*

5. Hello everyone, I think you have all read the instructions, you know well that there is a way allowing us to earn the maximum amount of money. Me and you, it’s a matter of trust, that’s why I simply invite you to choose... uh.... “Right”. Thank you for your confidence. *[Classification: non-promise]*

6. So... uh, I’m a competitor. I came here to play. Uh... so... uh, I’d like to have competitors in front of me. And... always choose an action, and... I will always roll my... my game. That’s it. *[Classification: promise]*

7. Hello everyone, uh, I think that in life sometimes you need to stick your neck out and... here, if you make that choice, you still have 5 chances out of 6 that it’s a good choice. Otherwise, you can decide to stay in your comfort zone, and make sure you get the minimum. And then, if everything is... if everything is a matter of luck, then maybe today you will make the right choice by trusting a stranger. That’s it. *[Classification: non-promise]*



8. Hello. So what I'm proposing is that the best solution is to cooperate. So I'm going to choose to roll the die every time. And if you can choose "Right", if you are adventurous and not afraid, the probability is still higher that you will win something. That's it. [*Classification: promise*]
9. Hello everyone, uh, so we're all going to play a game with stakes... monetary stakes. So you're player As, and I'm player B. So, um, you're going to choose between "Left" or "Right" and depending on that, so... uh... and I'm going to choose between rolling a die or not. So, uh, if you choose "Left", it's likely we'll all get the same gain. On the other hand, if you choose "Right", there are chances that I will not earn anything and that you will earn more. But... whatever you choose, I wish you all the best, and thank you. [*Classification: non-promise*]
10. Hello. Well, I think we're all here to make money, so I'm counting on you to be players. Don't take the easy way out, let's just leave it to chance, and then let's have some fun. That's it. Thank you very much. [*Classification: non-promise*]
11. Well, hello there. So my choice is easy, that is, either I earn a lot or I earn a little less but I help another person earn money. So it is a social ideal in the sense that I can create more wealth for more people, even if it is not for my benefit. You, your choice is more difficult. Of course, you can lose, while I have little to lose, and I have everything to gain from you taking the step of the social ideal. So I wish you good luck to think beyond individuality, and I look forward to seeing the results. [*Classification: promise*]
12. So this is the first time I'm taking part in an experiment like this. I didn't expect it to be about numbers. Uh, so it reminds me a little bit of the probability actually in... in high school, with math. Even though I haven't been doing it for three years now, it reminds me of that, actually. [*Classification: non-promise*]
13. Hello. I advise you, player As, to choose action "Right". If you take this decision, be aware that your gains from the experiment will be greater. If you choose the alternative action,

you will not have all the odds play in your favor. *[Classification: non-promise]*

14. Hello everyone. Uh, so trust and solidarity. If you are... if you trust me, I will act in solidarity. If you choose “Right”, I’ll roll the die, we’ll all be winners. So that’s it, you have to take risks. And with me, there’s no risk, if you go “Right”, I’ll roll the die. That’s it. *[Classification: promise]*

15. Well, I know it’s hard to trust someone, especially when you don’t know them. But this game is like in real life, it’s a kind of trusting and not trusting. So if you’re used to having... to trust people, well, bet on me. *[Classification: non-promise]*

16. So, I think... just like me, you have one chance in two to act as a player. And you have either... the individual choice, therefore... in the image of our society, if we put ourselves forward, we gain more at the expense of the other. And the other choice, which was to put us both, so player A and player B, on the same line, with positive effects for these two players. So I’ve made that choice. And others may have made the opposite choice. But, in any case, that’s it, I’m in... I’m not introducing myself because we’re not allowed to introduce ourselves, if I remember correctly. But here it is. *[Classification: promise]*

17. Well, hello to all of you. So you have 6 choices to make, so 6 chances to win. It’s sure that... choosing “Left” is safety, but choosing “Right” can make you earn twice as much. So, it’s up to you whether you’re a gambler or not. *[Classification: non-promise]*

18. Hello to all player As. In fact, we’re all here to increase our earnings, so we should help each other. So I propose you to choose the “Right” so that I... of course, I will choose “Roll”, so that, actually, we will have a chance of 5 out of 6 to have: you 12€ more, and me 10€ on top. So choose carefully, and above all think about how much you can earn. That’s it. *[Classification: promise]*

19. Listen, I won’t take long. If you choose “Left”, then we’ll have, no matter what I do, we’ll have 10€ to gain for two, meaning 5€ each. But if one takes “Right”... I see you as associates, and as associates we can get 22€ for two, 12 for you, and 10 for me. We have

a 80% chance... if we... and for the 20%, we have, uh, only 10€ for me, I guess, and you, you get nothing. We have a greater chance to gain together, well, individually, to gain more. Full stop. *[Classification: promise]*

20. Uh, hello. Actually, there is... I'm player B, I'm going to show you that... uh, you... you choose the... uh... Actually... because I forgot the thing, I don't know... Well, uh, I suggest you choose, choose "Don't roll". Because we, uh, it's not too beneficial compared to the... I don't know. *[Classification: non-promise]*

21. So as you know, you have the opportunity to... two possibilities, I meant. If you choose "Right", you have the possibility... your expected earnings can reach up to 12€. So you have more chances to win than by choosing "Left". If, on the other hand, you choose "Left", your expectation of gain is set at 5€. So your decision depends on whether or not you want to earn more money. Think carefully before making a decision because if you choose "Right", you can win up to 12€. If, on the other hand, you choose "Left", it is set at 5€. *[Classification: non-promise]*

22. Hello. In fact, in... we're in a game. And like in a game, we may or may not take risk. If you choose to take risk, you can earn even more, while if you do not take risk, the gains are certain, but... but they are low. OK. *[Classification: non-promise]*

23. Hello. Dear shareholders, I can reassure you that this action is a win-win situation. So don't think much and, uh, I suggest you go for it. *[Classification: non-promise]*

24. So, actually, for this experiment, I think that it's in the common interest that player B makes a decision, and I think rolling the die is a good decision. And player As, by choosing "Right", I think it's a good decision too, given that... By choosing "Left", there is a minimum gain, but by choosing "Right", there is more chance of getting a maximum gain, and it is more interesting and wiser to... to choose... uh... the possibility of earning a profit higher than what you could get at a minimum. So I think... it's wiser for player As to... to decide to choose the action "Right", and player B, well, to... to roll the die. *[Classification: promise]*

25. Well, choose “Right”. That way, you’ll make 12€ and I’ll make 10. It’s already better than 5€ each. If you played the lottery earlier, uh... trust me, you have a 5 in 6 chance of getting, uh, more than the double, while the lottery earlier was just double. And also like in the questionnaire, you have to trust others. *[Classification: promise]*
26. Hello to everyone, uh... we’re all here to make money. And above all, to win as much as possible. So, uh, I’m telling you, if you’re a player A, uh, you might as well choose “Right”, and if you’re a player B, you might as well choose to roll the die. That way, we’ll win between 10 and 12€ and... whereas if player A chooses “Left”, then we’ll only win 5€. That’s it. *[Classification: non-promise]*
27. Well, let me first say hi to everyone. I hope you’re enjoying yourselves. I guess you’ve all read the instructions, and that we all care about making some money. So, I think it’s better for you to choose “Right”, and for me to choose to roll the die. Because we have, well, we have 5 chances out of 6 that you get 12, and me only 10. I don’t mind. But if I sense you lack confidence in me, and that you’re tempted not to roll the die, then... no, to say “Left”, I will not roll the die. Which means we will all lose a lot. So make good choices. You know what to do. Choose “Right”, I will choose “Roll”. *[Classification: promise]*
28. Well, hello. What I wanted to tell you is: in this game, one shouldn’t, one shouldn’t be selfish or too "worried" because, as they say, too much thinking kills the game. One should know how to share the cake. Because, because in the end we’ll all get our... our pack of cigarettes. It’s just when you’re smoking your cigarette, you have to tell yourself that... how to say. You need to have a good conscience. Do you understand? So we have.... You know what the right decision to be made. And... All I have to say to you is: trust, follow your heart. Okay? Thank you. *[Classification: non-promise]*
29. Hello. In fact, from what I’ve seen, you have to, uh, actually, you choose the action that is most profitable for both of us, for both parties, and that will actually allow you to earn more than me. It’s okay. It’s okay. *[Classification: non-promise]*

30. Good morning, everyone. So, very simple: by choosing “Right”, you have one chance in six to lose 5€. But by choosing “Left”, you are sure to lose 7€. That’s it. [*Classification: non-promise*]
31. Anyways, let me explain. What is my option? It’s that you choose “Right” and I choose “Roll”. Anyways, we have.... 5 out of 6 chances that you will win more than me, and I – 1 out of 6 chances that I will win more than you, and you will lose. Because if it falls... if it falls 6, I’ll win 14 and you 0. And in the other cases, you 12 and I 10. So it’s better to... than to take 5€ / 5€. That’s my option. [*Classification: promise*]
32. So... well... now, I’m starting to bug here. Actually, I have a choice between rolling and not rolling, actually. In fact, I’m already going to... in fact, I’m already going to roll, it’s 100%. And you, you have a choice, you have a choice between “Right” and “Left”. “Right” and “Left”. Basically, if you choose “Right”... and then there will be... if I remember correctly... a die that will be rolled. If it falls between 1 and 5, you will... basically, you’re going to win 12 clams, I think. And if it falls... basically, on the number 6, you’ll lose everything. That’s it, it’s your choice. [*Classification: promise*]
33. Hello. Well, then we’ve all read the rules. We all know what it’s like, what is happening. All I have to say to you is that I’m going to cooperate anyway, so I’m going to roll the die because I don’t really care if I win 10 or 14. So if you can win 12, it’s good for everyone and everyone’s happy. No, I meant it! Believe me or not, you can do whatever you want. If you want to go home with five, that’s your problem. That’s it. [*Classification: promise*]
34. Hello, player As, I’m player B. So I invite you all to put “Right” each time, since you will lose a lot if you put something else. I’ll give you, I’ll give you... I commit myself to put “Roll” each time so that you can have a 1 in 5 chance of winning at least 12€. I hope I was convincing. [*Classification: promise*]
35. Hello to everyone. I have to make a speech. I don’t know what to say, but... there are two choices. There is “Left” or “Right”. If one takes “Left”, we’ll win 5€ each. If one takes

“Right” and I roll the die, we win... you win 12€ and I win 10€. I don’t want to make money that much, so if you want, you... you choose “Right” and I’ll roll the die, and... we’ll see, huh. I hope we win. *[Classification: promise]*

36. The goal of this game is to earn as much as possible. You have two choices. There is either the wise choice or... a safe choice. I would advise you on the wise choice, the second one. *[Classification: non-promise]*

37. Good morning, everyone. Well, I’m getting into this funny speech, with my face floured and... besides, I’m a little sick. I almost have a runny nose to, so it’s great. Mmm... yes, well, go ahead, be creative, go for it, or rather let me go for it and “Right”. That’s it. What else can I say? Mmm... yes... go ahead, be creative, that’s it. *[Classification: non-promise]*

38. Hello. Well, it’s going to be very simple. So I plan to choose to roll constantly. So I hope you will do the same, so that we can get the most out of it, that is, that you choose “Right”, in fact. That’s it. Thank you. *[Classification: promise]*

39. Hello everyone. So, in order for the roll to be taken into account and for us all have earnings, I advise you to do... to choose the option “Left”. That’s it. Thank you all. *[Classification: non-promise]*

40. Well, hello... Well, I’m going to choose to roll, so... if you’re a player... you have a 5 out of 6 chance of winning. That’s it. *[Classification: promise]*

41. Hello. Well... I’m not very good at speeches, and I didn’t really prepare anything, so... I’ll tell you that if we’re already here, let’s have fun! Have a good day. *[Classification: non-promise]*

## **F Who makes a promise?**

In Section 4.1, we have provided evidence that a promise to *Roll* is a credible strategic signal, since it is strongly and significantly predictive of the actual behavior. In this part, we take a step back

and address the following questions: *Who* makes a promise? Which individual characteristics can predict a player B making a promise to *Roll* in front of player As? To address this question, we follow the lead of a recent experimental study by Ismayilov and Potters (2016) who report (p. 390) that: *promises do not cause trustworthiness, they are just more likely to be sent by trustworthy players than by untrustworthy ones.*<sup>35</sup> Our experimental design allows us to screen the prospective promise-makers based on a rich set of individual characteristics elicited in the preliminary stage of the experiment and summarized in Table 5.<sup>36</sup>

Below, we show that pro-social distributional preferences (measured through the Social Value Orientation slider task by Murphy et al., 2011) are highly predictive of promise-making. Importantly, such preferences have been previously shown to explain trustworthy behavior in the trust game (Cox, 2004; Ashraf et al., 2006). Our result suggests that they also explain one’s willingness to make a promise which corroborates the conjecture made by Ismayilov and Potters (2016).

The econometric estimates of this screening exercise are reported in Table 7. For reasons stated in footnote 13, our econometric model accounts for session fixed effects. In our dataset, this precludes fitting a model using standard maximum likelihood method because of quasi-complete separation of observations.<sup>37</sup> Rather, we use two alternative specifications. First, we report coefficients of a linear probability model (LPM) with robust standard errors estimated through standard OLS procedure. As a robustness check, we report the outcomes of Firth’s logit model estimated through penalized maximum likelihood (ML) procedure.<sup>38</sup>

The main finding is that player Bs with a stronger pro-social orientation (as measured by their

---

<sup>35</sup>This conjecture differs at face value from the perspective commonly adapted by modern behavioral theories, like the lying aversion theory (Ellingsen and Johannesson, 2004), or the guilt aversion theory (Charness and Dufwenberg, 2006), that assume a causal link from promises to trustworthy behavior and explain its behavioral underpinnings.

<sup>36</sup>The distribution of the CRT scores is strongly left-skewed (see Appendix D for details) and thus the number of observations with a positive score is relatively small and quickly decreasing in the score. For this reason, in the regression analysis we use a binary variable  $1[CRT > 0]$  that splits the sample into two categories: those who correctly answered at least one question ( $1[CRT > 0] = 1$ ), and those who did not ( $1[CRT > 0] = 0$ ).

<sup>37</sup>In session 6, all player Bs made a promise to *Roll*, so that the corresponding indicator variable is a perfect predictor of “success”. This, in turn, makes it impossible to fit a binary response model through a standard ML exercise, since the value of the fixed effect for session 6 that maximizes the likelihood function is unbounded and tends to infinity.

<sup>38</sup>The Firth (1993) method has been proposed as a remedy for data separation problems (Heinze and Schemper, 2002). For Firth’s logit, it is recommended to use likelihood ratio (LR) test rather than the Wald test for significance testing (given that the estimated Hessian matrix does not allow for a reliable estimation of standard errors); for the same reason, odds ratios (OR) should be used instead of marginal effects.

Table 7: Individual predictors of promise-making

	LPM		Firth's logit		
	$\beta_{OLS}, (SE)$	$p$	$\beta_{Firth}$	OR	$p$
Intercept	0.214 (0.705)	0.764	0.178	1.195	0.976
SVO angle	0.014 (0.004)	0.001	0.120	1.128	0.010
1[ $CRT > 0$ ]	0.152 (0.172)	0.384	1.162	3.197	0.377
RMET score	0.026 (0.019)	0.182	0.302	1.353	0.096
GP investment	-0.004 (0.003)	0.185	-0.041	0.960	0.061
SOEP trust 1	0.064 (0.092)	0.494	2.201	9.035	0.443
SOEP trust 2	-0.072 (0.082)	0.389	-2.625	0.072	0.318
SOEP trust 3	-0.154 (0.144)	0.294	-2.973	0.051	0.194
Political views	0.007 (0.028)	0.810	0.203	1.225	0.570
Session FE's	YES		YES		
Joint insignificance tests					
Session FE's	0.001		0.029		
$\beta$ 's	<0.001		0.057		
$\beta$ 's and FE's	<0.001		0.060		

**Note.**  $N = 41$ . The first two columns provide OLS estimates ( $\beta_{OLS}$ ) of a linear probability model (heteroscedasticity-robust standard errors are given in parentheses) along with  $p$ -values from  $t$ -test for statistical significance.  $R^2 = 0.586$ . The second part of the table provides the results from an alternative specification, Firth's logit: estimated coefficients ( $\beta_{Firth}$ ), odds ratios (OR) and  $p$ -value from LR test for statistical significance. Joint significance tests in the lower part of the table are:  $F$ -test (LR test) for the left-hand (right-hand) side model.

SVO angle) are more likely to make a promise to *Roll*. The positive effect of pro-sociality on promise-making is highly significant and robust to the econometric specification ( $p_{LPM} = 0.001$ ;  $p_{Firth} = 0.010$ ).<sup>39</sup>

<sup>39</sup>Firth's logit provides some evidence that promise-making may also depend on emotional intelligence (OR increasing in the RMET score) and risk preferences (OR decreasing in the investment in the Gneezy-Potters task). However, these results are only (marginally) significant at the 10% level ( $p_{Firth} = 0.096$  and  $0.061$ , respectively), and they are not robust to the econometric specification of the model (respective coefficients are insignificant in the LPM with  $p_{LPM} = 0.182$  and  $0.185$ ).



Table 8: Robustness check: treatment comparisons controlling for player Bs fixed effects

	Diff. (SE)	$p$
VIDEO vs. AUDIO ( $H_0 : \alpha_1 - \alpha_2 = 0$ )	0.036 (0.042)	0.388
VIDEO vs. TEXT ( $H_0 : \alpha_1 - \alpha_3 = 0$ )	0.104 (0.041)	0.012
AUDIO vs. TEXT ( $H_0 : \alpha_2 - \alpha_3 = 0$ )	0.067 (0.039)	0.087

## G Robustness analysis for Figure 2

Table 8 re-assesses the comparisons reported in 1b after including player B fixed effects (41 indicator variables). Note that such robustness check is not meaningful for 1d, i.e. after including the promise/non-promise indicators which (unlike the treatment indicator variables) do vary within a given recording.

## H Trust and direct communication

Here, we compare player As' trust under direct (F2F) and indirect communication (VIDEO, AUDIO, and TEXT). Given that the experimental design of F2F differs from the remaining treatments in several ways, we would like to start by discussing the most important caveats for such comparison. Although we respond to each of them below, the results presented in the remaining part this section should be taken with a grain of salt.

One important concern is that the game is more “interactive” in F2F than elsewhere. In F2F, player A’s decision affects both players’ payoffs, while in the remaining treatments it does not affect player B’s gains (only the *vice versa* holds). As a result, as the role distributional concerns (“kindness”) may be more pronounced in player As’ decisions in F2F than elsewhere. Based on some of the previous evidence on distributive preferences and trust (see, e.g., Cox, 2004), one may expect this design asymmetry *per se* to stimulate trust in F2F relative to other treatments. We respond to (and allay) this concern in two ways. First, we do not observe such a pattern in our data: F2F

does not generate more trust than other conditions. We also note that so far, the experimental evidence is mixed regarding the role of distributional concerns in explaining trust. Cox (2004) estimates that distributional concerns are as important as the expected reciprocity. However, in a subsequent study Ashraf et al. (2006) find that kindness is not a primary driver of trust – most of the variance in trust is explained by expectations of return (which is a common factor in all our treatments). Second, in a study closely related to ours (in which players in a laboratory trust game communicate through pre-recorded, standardized video messages), Centorrino et al. (2015) collect data in two waves of experimental sessions. In the first wave ( $N = 84$ ), both players’ choices affected both players’ payoffs. The second wave ( $N = 114$ ) uses the same set of recordings, and this time player As’ decisions have no impact on player Bs’ gains. They do not find evidence that this design affects trust, that is, player As’ trust does not depend on the presence of monetary consequences for player Bs. <sup>40</sup>

Another design feature that distinguishes F2F from the other treatments is the number of decisions made by each player A (6 decisions in F2F and 10 decisions elsewhere). When parametrizing the VIDEO treatment, we also carried out a separate experimental session with 6 videos instead of 10. Clearly, the advantage of having 10 recordings per subject is an increase in the amount of experimental data. We did not find any evidence (either descriptive or statistical) that this increase in the number of observations per subject comes at a cost of affecting player As’ behavior.<sup>41</sup>

The overall trust rate in F2F is 40.2%. Descriptively, it is higher than in TEXT (36.5%), and lower than in AUDIO (44.4%) and VIDEO (49.0%); however, none of those differences is statistically significant (for statistical support, see Table 9). Like elsewhere, in F2F there is also a positive effect (equal to 10.2 percentage points,  $p = 0.099$ ) of player B’s promise to *Roll* on player A’s trust. Pairwise comparisons suggest that this effect is not significantly different from those observed in any of the remaining treatments; a joint  $F$ -test leads us to the same conclusion ( $p = 0.514$ ).

Altogether, these results suggest that making communication direct – i.e. moving from video

---

<sup>40</sup>Importantly, their experimental sample comes from a subject pool (university students in a large French city) that is quite similar to ours (university students in another large French city).

<sup>41</sup>We observed similar trust rates in the 6-round game, and in rounds 1-6 and 7-10 of VIDEO.

Table 9: Comparing F2F to other treatments: regression analysis

Frequency of:	<i>In</i>		<i>In</i>	
	$\lambda, (SE)$	<i>p</i>	$\lambda, (SE)$	<i>p</i>
Intercept ( $\lambda_0$ )	0.402 (0.047)	< 0.001	0.350 (0.055)	< 0.001
1[ <i>VIDEO</i> ] ( $\lambda_1$ )	0.088 (0.058)	0.128	0.070 (0.068)	0.309
1[ <i>AUDIO</i> ] ( $\lambda_2$ )	0.042 (0.055)	0.451	-0.008 (0.065)	0.908
1[ <i>TEXT</i> ] ( $\lambda_3$ )	-0.038 (0.054)	0.488	-0.049 (0.063)	0.434
1[ <i>Promise</i> ] ( $\lambda_4$ )			0.102 (0.062)	0.099
1[ <i>Promise</i> ] $\times$ 1[ <i>VIDEO</i> ] ( $\lambda_5$ )			0.029 (0.071)	0.686
1[ <i>Promise</i> ] $\times$ 1[ <i>AUDIO</i> ] ( $\lambda_6$ )			0.090 (0.072)	0.214
1[ <i>Promise</i> ] $\times$ 1[ <i>TEXT</i> ] ( $\lambda_7$ )			0.031 (0.072)	0.671
	Effect of promise on trust rates across treatments			
F2F ( $H_0 : \lambda_4 = 0$ )				0.099
VIDEO ( $H_0 : \lambda_4 + \lambda_5 = 0$ )				< 0.001
AUDIO ( $H_0 : \lambda_4 + \lambda_6 = 0$ )				< 0.001
TEXT ( $H_0 : \lambda_4 + \lambda_7 = 0$ )				< 0.001
Joint <i>F</i> -test ( $H_0 : \lambda_5 = \lambda_6 = \lambda_7 = 0$ )				0.514

**Note.** Estimates of linear probability models (LPM). Dependent variable: indicator of player A's decision to go *In*. In the first model, the dependent variable is regressed on the all-experiment set of treatment dummies (= 1 for a given treatment, = 0 otherwise). Each of the estimated coefficients corresponds to the average effect of the underlying experimental condition on the probability of observing a given outcome, relative to the reference F2F treatment (the outcome of which is captured by the intercept). The second model also include a dummy variable 1[*Promise*] (= 1 if a player A received a promise to *Roll* from player B, = 0 otherwise), and its interactions with each of the treatment indicator variables. Models use cluster-robust standard errors (with session-level clustering for the F2F treatment and individual-level clustering in the remaining treatments) to account for the potential unobserved correlation between individual observations, as well as the heteroscedasticity of residuals in the LPM. There are 246 observations (7 clusters) in F2F, 620 (62) in VIDEO, 700 (70) in AUDIO, and 850 (85) in TEXT.

messages to face-to-face interaction – does not improve trust. If anything, it may be counterproductive (although the differences between those two conditions are not statistically significant), and the outcomes from the face-to-face communication treatment are very close to those in the baseline treatment with plain text messages. This is in line with the previous results from public goods games and trust games reported in Bochet et al. (2006); Bicchieri et al. (2010), but goes against the findings in Balliet (2010) who suggests that face-to-face communication is a far more powerful device for inducing cooperation than written messages.

# I Player As' perceptions of player Bs' characteristics

Table 10 summarizes Player A's' perception of player Bs in our experiments.

Table 10: Player As' perceptions of player Bs on a 1-8 scale

Average score on:	beauty	intelligence	trustworthiness	self-confidence
	F2F			
No promise	4.817	5.150	4.367	4.608
Promise	4.921	5.421	4.484	5.071
Ranksum $p$	0.754	0.289	0.556	0.327
Overall mean score	4.870	5.289	4.427	4.846
	VIDEO			
No promise	4.661	5.112	4.476	4.493
Promise	4.704	5.611	4.760	5.198
Ranksum $p$	0.876	0.015	0.167	0.074
Overall mean score	4.684	5.381	4.629	4.873
	AUDIO			
No promise	4.706	4.921	4.291	4.973
Promise	4.878	5.476	4.557	5.757
Ranksum $p$	0.256	0.020	0.175	0.037
Overall mean score	4.797	5.214	4.431	5.387
	TEXT			
No promise		4.971	4.145	5.290
Promise		5.407	4.412	5.623
Ranksum $p$		0.124	0.179	0.322
Overall mean score		5.180	4.273	5.449

**Note.**  $p$ -values come from ranksum test using a player B's average score in player As' evaluations as a single observation. Comparisons are made between promise-makers (21 observations) and others (20 observations).