

Beyond Words: Non-Verbal Cues in Virtual Collaborations*

Michelle Hörrmann, Petra Nieken, Sven Walther

This is a preliminary version. Please do not cite or distribute.

Abstract

Virtual communication has become commonplace, especially for geographically dispersed teams collaborating on joint projects despite lacking prior relationships. These teams will likely encounter coordination failures, resulting in low overall cooperation rates. However, non-verbal cues transmitted during communication might help to mitigate this problem. Our study investigates whether and how non-verbal cues in virtual team meetings can foster collaboration. We implemented a weakest-link game in a controlled large-scale online experiment on Prolific. Before playing the weakest-link game, team members participated in a non-game-related virtual team meeting to get acquainted. We systematically varied the availability and amount of non-verbal cues between our treatments during this meeting phase. We expected these cues to be crucial in overcoming strategic uncertainty and fostering group cohesion. Depending on the treatment, the virtual team meeting was an audio-only meeting, an audio meeting with selfies, or a video meeting. We observe significantly higher cooperation in the audio-only and the video meeting than in the audio meeting enriched with selfies. Static non-verbal cues, thus, had a negative effect on team cooperation. We find that the audio-only and the video meeting led to similar team cooperation and that coordination did not differ between treatments. Our findings shed light on how virtual teams can enhance their collaboration by leveraging the power of non-verbal cues. By understanding the impact of non-verbal cues, organizations can foster stronger connections and achieve more successful outcomes.

*We would like to thank John Antonakis, Brice Corgnet, Adam Zylbersztejn, the audience at the ESA North American Meeting 2024, the GfEW Annual Meeting 2024, the KD²School Autumn Retreat 2024, the 1st European Economic Review Summer School in Experimental and Behavioral Economics, and KIT's Brown Bag Seminar in the summer term 2024 for helpful comments and suggestions. We thank Sergiu Panainte, Roxanne Weber, and Barbara Sammari for their excellent research assistance. (This project was funded as part of the Excellence Strategy of the German Federal and State Governments (Future Fields 3 - Decision-Making in Hybrid Adaptive Systems for Better Work and Life – An Open Science Approach) and received financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through GRK2739/1 (Project Nr. 447089431 – Research Training Group: KD²School – Designing Adaptive Systems for Economic Decisions)).

Keywords: Communication Channels, Weakest-Link, Cooperation, Virtual Collaboration

JEL-Codes: C99; D83; D91; M50

1 Introduction

‘Please turn on your camera during the meeting’ is a common phrase at the start of virtual meetings to encourage engagement and connection via video. 92 percent of U.S. executives even go as far as stating that employees who frequently turn off their cameras (or mute themselves) will not have a long-term future with the company (Vyopta, 2022), but still, many virtual meetings keep relying on audio and therefore verbal and para-verbal cues only.

Since the Covid-19 pandemic, virtual meetings have become common for many at work or in their daily lives and are expected to persist in the future (Barrero et al., 2023; Tolliver and Sass, 2024). While virtual meetings facilitate collaboration among geographically dispersed teams, they are also increasingly used in hybrid work arrangements, even when employees are co-located on office days (Bloom et al., 2022). Together with the ongoing globalization, this leads to new teams meeting virtually for the first time. Therefore, it is important to understand how enriching virtual pre-play communication with non-verbal cues affects team collaboration.

Encouraging new teams to turn on their cameras can have several underlying reasons. On the one hand, it enables teams to gain a lot more information about others through the added non-verbal cues as explained by the media richness theory (Daft and Lengel, 1986). The availability of non-verbal cues simultaneously makes the communication more similar to well-known face-to-face meetings and provides all meeting participants with more information to judge their team members better, e.g., to gain an understanding of their intentions and their type (Eckel and Petrie, 2011; He et al., 2017). Furthermore, pre-play video communication has been shown to lead to significantly higher levels of cooperation compared to audio communication and to levels similar to those in face-to-face meetings (Brosig and Weimann, 2003), proposing video meetings as a good substitute for in-person meetings.

However, an increase in non-verbal cues through a video can also have negative impacts on meetings, such as non-verbal overload, leading to zoom fatigue or less social connectedness of meeting participants (Fauville et al., 2023; Queiroz et al., 2023). Thus, another real-world example of how to enrich virtual meetings with non-verbal cues is the use of static photos similar to profile pictures in MS Teams or Zoom. This could lead to a reduction in overload while simultaneously allowing for the visual identification of others. Being able to identify others, such as by name or photo, reduces

anonymity, is very useful for promoting prosocial behavior (Andreoni and Petrie, 2004; Charness and Gneezy, 2008), and has an effect on trust and efficiency (Eckel and Petrie, 2011; Bohnet and Frey, 1999).

To our knowledge, it has not been investigated yet whether adding static photos to audio communication leads to similar results as either audio or video meetings on team cooperation and coordination. In addition, the effect of communication on coordination and cooperation has previously usually been studied with game-related communication including promises and commitment (see e.g., Brosig and Weimann (2003); He et al. (2017); Avoyan and Ramos (2023)), while we add to it by having a communication stage that is not game-related to focus on team building and the effect of non-verbal cues.

This study intends to fill these gaps by addressing the main research question: Does enriching virtual pre-play communication with non-verbal cues increase team coordination and cooperation? We conducted a controlled online experiment in two parts. While the treatment-independent Part 1 was mainly used to recruit subjects for Part 2 and to collect selfies of subjects, the main experimental manipulation took place in Part 2. Part 2 was completed in fixed teams of three subjects and consisted of two main stages: a communication and a team decision stage. In the communication stage, subjects took part in a virtual meeting, the format of which varied depending on the treatment. In the baseline (*Audio* treatment) with no non-verbal cues, subjects could only hear each other, in the *Photo* treatment they could additionally see selfies of all team members', and in the *Video* treatment they saw the live videos of each other. The main change between the *Audio* and *Photo* treatment was thereby the possibility of visual identification through static non-verbal cues (selfies). From the *Photo* to the *Video* treatment, turned-on cameras further increased the amount of available non-verbal cues by adding dynamic components, such as facial expressions and gestures. Importantly, the communication stage was the only treatment-dependent stage and not game-related, as subjects only received the instructions for the team decision task afterwards. In the team decision stage subjects played the weakest-link game (Van Huyck et al., 1990; Brandts and Cooper, 2006a) with limited feedback for ten rounds. Note that only decisions in Round 1 were relevant to our hypothesis as they capture immediate treatment differences.

The data show that enriching virtual pre-play communication to transmit more non-verbal cues had no significant effect on coordination in the first round of the weakest-link game. For cooper-

ation, however, we identified significant treatment differences: (i) Adding static non-verbal cues with a selfie marginally significantly decreased team cooperation compared to the baseline *Audio*, and (ii) compared to a static selfie, live interaction via video significantly increased cooperation. Nevertheless, there was no difference in cooperation between the *Audio* and *Video* treatment.

Our paper contributes to three areas of research: (i) Research on team coordination and cooperation, in particular in virtual settings, (ii) literature on different communication channels, and (iii) literature relating to identification, anonymity, and social distance.

The positive effect of communication on economic behavior and decision-making has already been demonstrated in general (Cooper et al., 1992; Crawford, 1998; Charness and Dufwenberg, 2006; He et al., 2017), as well as its role in enhancing coordination and cooperation, particularly in the weakest-link game (Chen and Chen, 2011; Dong et al., 2018; Toku et al., 2022; Avoyan and Ramos, 2023). Our study contributes to that literature by providing an in-depth analysis of the role of non-verbal communication. A main difference between our study and the existing literature is that communication was not game-related and, therefore, did not allow for promises and commitments but instead focused on the non-verbal parts of communication which had not yet been investigated in this context. The results do not show any positive effects on cooperation or coordination when adding static or dynamic non-verbal cues to an audio-only meeting. In contrast, static non-verbal cues in the form of selfies, which provide visual identification, decrease team cooperation compared to both no non-verbal cues and dynamic non-verbal cues in the form of a live video. Similarly, social identity (Chen and Chen, 2011) and group cohesion (Gächter et al., 2024) had both been studied in relation to coordination and cooperation. Our results extend previous research that found beliefs and group cohesion to be crucial for promoting cooperation (Gächter et al., 2024). While Gächter et al. (2024) manipulated group cohesion exogenously by putting either friends or strangers in groups, we show that varying the availability of nonverbal cues in virtual communication can also significantly influence the emergence of cohesion within groups.

Furthermore, our findings contribute to the literature studying virtual communication with a focus on how different communication channels affect economic behavior such as charisma and performance (Nieken, 2023), creativity (Grözing et al., 2020), trust and trustworthiness (Zylbersztein et al., 2020; Babutsidze et al., 2021; Zylbersztein et al., 2021), affiliation with communication partners (Sprecher, 2014), or charitable giving (Zylbersztein et al., 2024b). More related to our paper,

Brosig and Weimann (2003), Bochet et al. (2006), and Frohlich and Oppenheimer (1998) studied the effect of different communication channels on cooperation. While the study by Brosig and Weimann (2003) with game-related communication revealed a significantly higher rate of cooperation when using video compared to audio communication, we found no evidence that this is also the case for non-game-related communication. Furthermore, our findings contribute to the understanding of how different degrees of human presence emerge in virtual communication, which is particularly relevant for reducing dishonest behavior (Cohn et al., 2022; Nieken and Walther, 2024).

Our *Photo* treatment allowed us to isolate the effect of visual identification and, thus, our paper links to an extensive body of research related to identification, anonymity, and social distance. While other studies found positive effects of identification (e.g., Andreoni and Petrie (2004); Charness and Gneezy (2008); Eckel and Petrie (2011), visual identification did not appear to have the same effect in our study. Our results revealed that enriching virtual communication with more non-verbal cues by adding selfies of the subjects or turning on video during the communication significantly reduced the perceived sense of anonymity. However, anonymity was not of relevance for team behavior in terms of coordination and collaboration in any treatment. Instead, human presence, which is significantly lower at the 10% level with static non-verbal cues, indirectly affected team cooperation.

The remainder of the paper is organized as follows. Section 2 presents our study design, as well as our treatments and hypothesis. In section 3, we give an explanation of our main variables, our sample, and empirical strategy. We describe and discuss our results in the following sections 4 and 5 and then conclude with section 6.

2 Experimental Design and Hypotheses

In this section, we present our general study design, our procedure as well as our treatments and hypotheses.

2.1 General Description

We conducted a controlled online experiment on Prolific, consisting of two parts which were conducted with several days difference.¹ Part 1 was elicited with SoSci Survey (Leiner, 2024), while in Part 2 we used the experimental software oTree (Chen et al., 2016). Subjects were recruited in the first part of the study, where they completed a brief survey and took a selfie as a necessary preparation for the second part. Part 2 then contained the main experimental manipulation, with three subjects randomly forming a team. Each team went through two main stages. First, a communication stage, in which they met their team members in a virtual meeting,² and thereafter, a team decision stage. The teams played a weakest-link game for ten rounds to measure team coordination and cooperation. Lastly, subjects individually answered a brief survey.

2.2 Study Part 1

In Part 1, subjects were recruited via Prolific, and all subjects were located in the United States.³ This part of the study was identical for all treatments and served two main purposes: (i) collect subjects’ selfies as well as audio recordings, and (ii) select study time slots for participation in Part 2. First, the subjects received general instructions and made some technical checks to test their audio and video. Then, subjects took a selfie and an audio recording of their voice with a pre-specified sentence.⁴ These were necessary for the *Photo* treatment as well as for identification in Part 2 of the study. Since some Part 2 questions had to be answered for each team member individually, we provided subjects with the audio recordings and, in the *Photo* and *Video* treatment, with all team members’ selfies to facilitate identification. Next, subjects got a list of possible time

¹The study has been pre-registered at OSF and approved by the ethics committee of the Karlsruhe Institute of Technology.

²Virtual meetings were facilitated using the open-source platform Jitsi (<https://jitsi.github.io/handbook/>) integrated into oTree via its iFrame API. We used Jitsi for two reasons. First, we were able to host it locally on servers hosted by our university to comply with data protection rules. Second, the setting allowed us to customize the meeting software to our requirements using the iFrame API (e.g., turning off the camera function or integrating subjects’ selfies).

³In addition, we used Prolific prescreeners for age (18 to 65), fluent languages (English), approval rate (min. 95%), number of previous submissions (min. 100), and record video (Yes). Due to the age restriction and the language screener, we had a sample of subjects who were of working age and had no language barrier during the virtual meeting. In addition, they must have previously agreed to record a video of themselves to reduce return rates because they had to upload a selfie, and we recorded the audio during the communication stage. As the resulting Prolific sub-sample was female-dominated (as of September 2024), we used a quota sample on gender to ensure similar numbers of male and female subjects in our study.

⁴“This is a test for the virtual team communication study.”

slots to participate in Part 2 and answered a brief survey.⁵ The average duration of Part 1 was about 11.2 minutes, and subjects received a fixed payment of £2.00 independent of whether they participated in Part 2. This results in subjects’ average hourly earnings of £10.81 in Part 1.

Study Part 1 and Part 2 were always held on different days. The advantage of splitting our study into two parts was that we could elicit specific measures before the treatment manipulation to eliminate some potential endogeneity problems. In particular, the time gap between Part 1 and 2 should mitigate potential influences of these measures on the weakest-link game outcomes. In addition, we needed to review subjects’ selfies and audio recordings before Part 2 to exclude subjects that did not complete Part 1 successfully (e.g., if subjects’ faces were not displayed in the selfies or subjects did not record the pre-specified sentence). Furthermore, the backgrounds of the uploaded selfies were edited before Part 2. We grayed out the photo backgrounds to exclude any influences from the subject’s surroundings.

2.3 Study Part 2

The second part of the study consisted of two main stages: (i) a communication, and (ii) a team decision stage. First, we describe the treatment manipulation, the weakest-link game and the team decision stage, before providing a detailed description of the Part 2 procedure.

2.3.1 Treatments

Our experiment encompassed three treatments (*Audio*, *Photo* and *Video*). The treatments only differed in the communication stage, i.e., the virtual meeting.⁶ In all treatments, subjects were able to communicate via audio. The treatments differed in the amount and type of non-verbal cues that were available in the virtual meeting. The *Audio* treatment served as our baseline compared to the *Photo* and *Video* treatment. Subjects could only hear each other as cameras were exogenously turned off. Therefore, only verbal and para-verbal cues were possible, but no non-verbal cues.

⁵This survey captured demographic data, the Inclusion of Other in the Self Scale (IOS) (Aron et al., 1992), the We Scale (Cialdini et al., 1997), risk and social preferences (reciprocity, altruism) (Falk et al., 2023), interpersonal trust (Nießen et al., 2020), and collective orientation (Driskell et al., 2010; Hagemann, 2020).

⁶The identification of individual team members in the questionnaires posed an exception. In the *Photo* and *Video* treatment, the selfie and audio recording from Part 1 were shown with the subject’s generic ID, while only the audio recording was provided in the *Audio* treatment. A second exception was the attractiveness rating, in which, depending on the treatment, only the attractiveness of the voice (*Audio*) or also the physical attractiveness of the other team members (*Photo* and *Video*) was captured.

In the *Photo* treatment, cameras were also exogenously turned off.⁷ Compared to the *Audio* treatment, in which subjects could only see images of subjects’ generic IDs (“team member E”, “team member O”, “team member I”) (see Figure 1), in the *Photo* treatment we added static non-verbal cues. This was done by replacing the images of generic IDs with subjects’ selfies that they took in Part 1 of the study. Therefore, the *Photo* treatment eliminated visual anonymity through the inclusion of static non-verbal facial cues.

In the *Video* treatment, we included dynamic non-verbal cues. Subjects’ cameras were turned on, and thus, they could see themselves and their team members. This live video interaction further increased the amount of non-verbal cues, e.g., mimics, gestures, and posture.

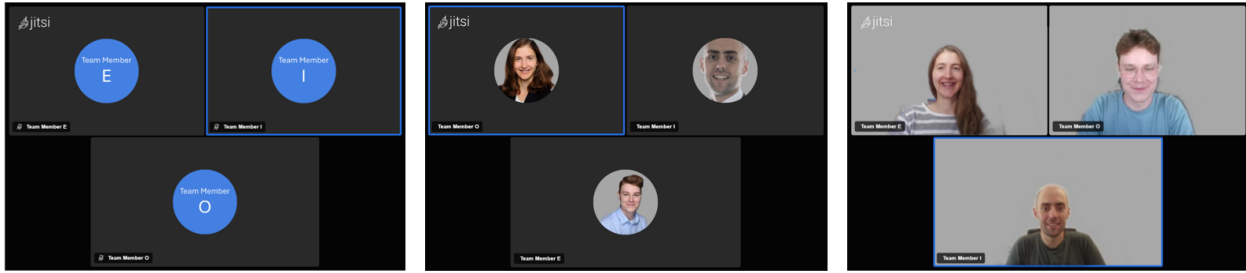


Figure 1: Virtual meeting setup (from left to right): *Audio*, *Photo*, *Video*

2.3.2 Weakest-link Game

In the team decision stage, we used the weakest-link game, which is a standard economic game allowing us to measure both coordination and cooperation (see Subsection 3.1 for details). We chose the naturalistic weakest-link game framing by Brandts and Cooper (2006a) in which team members acted as employees and were asked to allocate work hours (the term used for effort levels) between an individual project P_B and a team project P_A . In each round, subjects had to choose one of the following effort levels h_A to allocate to the team project: $h_A = \{0, 10, 20, 30, 40\}$.

The payoff in the weakest-link game depended on the own effort h_A as well as the minimum effort in the team h_{min} . The own effort was costly with $c = 5$ leading to costs of $5 \times h_A$ and the minimum effort was rewarded with a bonus. We set $B = 10$ which is one of the bonus rates used by Brandts and Cooper (2006a). The payoff function in our setting, therefore, was as follows:

$$\pi = 200 - 5 \times h_A + 10 \times h_{min}$$

⁷In the *Audio* and *Photo* treatment, subjects were not able to turn on the camera function in Jitsi.

Each time subjects made a decision in the weakest-link game, they saw the resulting payoff table (see Figure 2), which showed their possible payoffs for all combinations of their own possible effort and the possible minimum effort of all other team members.

Payoff table	Minimum number of hours contributed to Project A by any team member other than yourself					
		0	10	20	30	40
Your number of hours contributed to Project A (h_A)	0	200	200	200	200	200
	10	150	250	250	250	250
	20	100	200	300	300	300
	30	50	150	250	350	350
	40	0	100	200	300	400

Figure 2: Payoff Table for the Weakest-Link Game

Following Dong et al. (2018), coordination on the highest effort level is expected if $n < B/c$ and coordination on the lowest effort level for $n > B/c$. Therefore, for the parameters $n = 3$, $c = 5$, and $B = 10$, coordination on the lowest effort level was expected.

The resulting weakest-link game had five pareto-ranked pure Nash equilibria which were the strategy profiles in which all team members chose the same effort level. The risk-dominant equilibrium in which all team members chose an effort of zero was at the same time ranked lowest in terms of pareto efficiency. Contrarily, all team members choosing the highest effort level led to the pareto- and payoff-dominant equilibrium (Riedl et al., 2016; Van Huyck et al., 1990).

2.3.3 Procedure

Part 2 started with information on the different study stages and the payoff. Next, subjects had to complete some technical checks and affirm that the selfies and audio recording they saw were the ones they uploaded in Part 1.

As stated in Section 2.1, in the communication stage, teams of three subjects had the opportunity to communicate in a virtual meeting. Subjects were grouped into these teams based on their arrival time and then received instructions for the virtual meeting. Immediately before the virtual meeting, subjects received instructions for a brief icebreaker task following Kaiser et al. (2022) and Koudenburg et al. (2013). They individually rated different types of vacations⁸ on a 7-point Likert scale (see Figure 5 in the Appendix). Next, we asked them to discuss these vacation types

⁸sun, sea and beach vacation; party vacation; winter sports vacation; city vacation; educational vacation; camping vacation; cruise vacation

during the virtual meeting, a method which has been established as an icebreaker to promote conversational flow (Kaiser et al., 2022; Koudenburg et al., 2013). In addition, it allowed us to control for similarities within teams. As soon as all team members completed this task, the virtual meeting started. Beyond that, subjects received no further instructions regarding the content of the communication and were free to communicate as they wanted. However, it is important to note that subjects did not receive any information about the task in the subsequent stage, i.e., the weakest-link game. Subjects were only informed that they would, after the meeting, be grouped again with the same team. Thus, the communication content was not game-related. Communication was used in our study to get to know other team members, and not for promises, or to discuss other content related to the game. The virtual meeting was the only treatment-dependent stage of the experiment. It entailed seven minutes of unrestricted communication, and ended automatically. Since we grayed out the photo backgrounds for the *Photo* treatment to exclude any influences from the subject’s surroundings, in the *Video* treatment, we likewise asked subjects to use a gray background. After a successful virtual meeting (see 3.2 for our exclusion criteria), the teams proceeded with the team decision stage.

At this point, subjects received the instructions for the weakest-link game and played ten rounds in their fixed teams of three. After the first and the last round of the weakest-link game, we elicited subjects’ beliefs about the effort levels of their other two team members. After every round of the weakest-link game, subjects received limited feedback (Brandts and Cooper, 2006b) about the minimum effort in the previous round but not about the individual effort of their team members. After the ten rounds of the weakest-link game, subjects individually answered a brief survey to elicit team-specific information. We used the IOS scale (Aron et al., 1992), the We scale (Cialdini et al., 1997) and an attractiveness rating⁹ (He et al., 2017) for each team member. In addition, perceived human presence (Cohn et al., 2022) and perceived anonymity (Hite et al., 2014) were elicited. To conclude the study, subjects were asked if they knew any of the other team members prior to the meeting, and in the *Video* treatment if the other team members had selected the gray background during the meeting.

As Part 2 was carried out mainly in a team and dropouts and waiting times could be problematic, we followed the recommendations of Arechar et al. (2018). As soon as subjects were grouped at

⁹In the *Audio* treatment subjects rated attractiveness of the voice only, in the *Photo* and *Video* treatment physical attractiveness was rated additionally.

the beginning of the communication stage, each study page was equipped with on-screen timers. If subjects did not complete a page in time, they were counted as a dropout, and their team members were notified. If a dropout occurred at any point during the communication stage, the other team members could not complete the study and were paid in part, compensating them for their time. Dropouts during the team decision stage did not immediately terminate the study for the whole team. Instead, the remaining team members could, with the knowledge that a team member dropped out, continue with the weakest-link game,¹⁰ finished the study, and were paid in full.

Subjects received a fixed payment of £6.00 for participation and a bonus payment. This bonus payment depended on the outcome of the weakest-link game, and a computer randomly selected one of the ten rounds (with equal probability to mute potential income effects) as payoff-relevant. The subjects received, on average, 240.44 ECU as a bonus, which corresponds to an average of £4.81. In total, this results in subjects' average earnings of £10.81 in Part 2. Part 2 had an average duration of about 32.1 minutes, which results in average hourly earnings for subjects of around £20.17.

2.4 Hypotheses

The main objective of our paper is to study how different non-verbal cues impact cooperation and coordination. As we are interested in the immediate treatment effects on coordination and cooperation, we focus our hypotheses on the first round of the weakest-link game. As the weakest-link game was played in a limited feedback setting (Brandts and Cooper, 2006b), subjects received information on the team minimum after each round. Therefore, treatment effects might be dampened after the first round.

Recall that in the *Audio* treatment there were no non-verbal cues. Therefore, this treatment served as a baseline compared to the *Photo* and *Video* treatment. In the *Photo* treatment, subjects received additional static non-verbal cues, as they could see team members' selfies. In the *Video* treatment, we increased the amount of non-verbal cues in the virtual meeting even further. In addition to team members faces, subjects received dynamic non-verbal cues in a virtual video meeting

¹⁰Only the decisions of the remaining team members were relevant to determine the payoff, and this was communicated to the remaining subjects.

with turned on cameras.

Based on literature on communication channels (e.g., Bochet et al. (2006); Brosig and Weimann (2003)), group cohesion (Gächter et al., 2024), identification (Andreoni and Petrie, 2004; Bohnet and Frey, 1999; Eckel and Petrie, 2011), and media richness (Daft and Lengel, 1986; Balliet, 2010), we expected that increasing the amount of non-verbal cues in a virtual meeting would have positive effects on coordination and cooperation. In particular, in the *Photo* treatment compared to the *Audio* treatment, visual anonymity was reduced because visual identification of team members was added. The identification of others has been shown to increase cooperation in a public goods game (Andreoni and Petrie, 2004) and in a prisoner’s dilemma (Bohnet and Frey, 1999), enhance trustworthiness in trust games (Eckel and Petrie, 2011) and to facilitate the detection of whether responders accepted or rejected offers in ultimatum games (Van Leeuwen et al., 2018). When comparing the *Video* treatment with the other two treatments, following the media richness theory from Daft and Lengel (1986), we expected that video would lead to higher coordination and cooperation. In a public goods game video communication significantly increased cooperation compared to audio communication (Brosig and Weimann, 2003) and non-verbal cues, such as smiles, winks or handshakes, that were in our setting only available in the *Video* treatment, facilitated group cooperation in a weakest-link game (Manzini et al., 2009).

This leads us to the following hypotheses:

H1: Enriching virtual pre-play communication to transmit more non-verbal cues leads to higher levels of coordination in teams in the first round of a weakest-link game.

H2: Enriching virtual pre-play communication to transmit more non-verbal cues leads to higher levels of cooperation in teams in the first round of a weakest-link game.

In particular, we expect that (a) coordination/cooperation in the *Photo* treatment is higher than in the *Audio* treatment, and (b) coordination/cooperation in the *Video* treatment is higher than in the *Photo* treatment.

3 Data and Estimation Strategy

In the following, we describe the main variables used in the analyses and provide information on the sample and our estimation strategy.

3.1 Variables of Interest

To test our hypotheses, we use two main variables to measure coordination and cooperation. These variables are used based on Gächter et al. (2024), who define coordination as matching the effort level of other team members (reaching any Nash equilibrium) and cooperation as achieving pareto-superior equilibria. Therefore, if coordination is high, subjects will select the same effort level, resulting in a low wasted effort. Besides, high cooperation requires a high minimum effort in the team. In line with this definition, we choose the *Team Minimum Effort in round r* and the *Team Wasted Effort in round r* as our main variables.

Team Wasted Effort in round r

To test our first hypothesis (coordination), we use the wasted effort in round $r = 1$ on the team-level. The team's wasted effort is the total sum of the individual efforts in a team minus the total sum of the minimum efforts in a team: $Team\ Wasted\ Effort = \sum_{i=1}^n (x_i - x_{min})$.

Team Minimum Effort in round r

Similarly, to test our second hypothesis (cooperation) we use the minimum effort in round $r = 1$ on the team-level: $Team\ Minimum\ Effort = \min_i(x_i)$. As subjects select from the possible effort choices $\{0, 10, 20, 30, 40\}$, the *Team Minimum Effort* in round 1 will also be one of these values.

We use the indicator variable *Photo* (*Video*) to analyze pairwise differences between our treatments, which is one if a subject or team was part of the *Photo* (*Video*) treatment and zero otherwise.

3.2 Sample Selection

In total, 1775 subjects completed Part 1. We excluded 12 subjects because they failed both attention checks. In addition, we excluded 89 subjects because their uploaded selfie or audio recording did not fulfill the requirements explained in the upload instructions (e.g., subject's selfie did not display their face or did not record their voice). As a result, 1674 subjects who met our requirements were invited to participate in Part 2. From these subjects, 963 fulfilled our requirements to start the communication stage in Part 2 and were grouped into teams of three.¹¹ Thus, 321 teams were able to start the communication stage, 105 teams in the *Audio* treatment, 93 teams in the *Photo* treatment, and 123 in the *Video* treatment. Of these 321 teams, 204 teams were able to finish the study, forming our analytical sample (67 teams in the *Audio* treatment, 68 teams in the *Photo* treatment, and 69 teams in the *Video* treatment).¹² All teams with technical issues

¹¹Before the communication stage, subjects had to pass a browser, audio, and video check and had to confirm their uploaded photo and audio recording from Part 1. Furthermore, we could not form a team of three for some subjects because not enough subjects were available during the time slot.

¹²Despite our request for subjects to use a gray background in the *Video* treatment, in none of the teams all three members used the gray background. As it was technically not possible to exogenously set the video background, we could not use this as an exclusion criterion.

during the virtual meeting or dropouts were directly excluded from the study or excluded from our analytical sample.

3.3 Empirical Strategy

We started the analysis of our hypotheses with non-parametric tests. First, we investigated the differences between all three treatments and, subsequently, examined all pairwise differences. Then, we estimated a series of linear regressions where i indexes teams.

To investigate Hypothesis 1, we analyzed the teams' coordination behavior and first introduced the variables *Photo* and *Video* to show the pure treatment effects. In the second specification, we added the control variables *Homophily*, *Team Cohesion Part 1*, and *Team Risk Aversion*. *Homophily* refers to an index that increases with the similarity of team members in terms of age, gender, and ethnicity. *Team Cohesion Part 1* and *Team Risk Aversion* control for team average preferences on team cohesion and risk aversion that we elicited before the communication stage in Part 1 of our study. We follow the same strategy to investigate Hypothesis 2. This results in the following regression equation:

$$Y_i = \beta_0 + \beta_1 photo_i + \beta_2 video_i + \beta_3 controls_i + \epsilon_i$$

Y refers to the *Team Wasted Effort* and *Team Minimum* in round 1.

4 Results

In the following, we first analyze coordination in the weakest-link game and, afterward, cooperation. As pre-registered, our main analysis covers non-parametric tests and regressions investigating coordination and cooperation in Round 1, then, we conduct some exploratory analyses to investigate individual behavior, behavior over rounds, and potential mechanisms.¹³

4.1 Coordination

We start with descriptive information about the teams' coordination behavior in Round 1 in all three treatments. Recall that we measure coordination with *Team Wasted Effort*. Thus, higher levels of coordination are measured by lower levels of *Team Wasted Effort*. In the baseline (*Audio* treatment) with no non-verbal

¹³Randomization checks showed no significant differences in demographics (age, gender, ethnicity, education) and social preferences (in particular *Team Risk Aversion*) between the treatments. Overall, *Team Cohesion Part 1* scores were also not significantly different between treatments (Kruskal-Wallis test, $p = 0.1457$). Only a pairwise comparison between the *Photo* and *Video* treatment revealed a significant difference in *Team Cohesion Part 1* at the 10% level (Mann-Whitney U test, $p = 0.069$).

cues, *Team Wasted Effort* was, on average, 24.33 (SD = 15.20). We observe higher levels of 28.38 in the *Photo* treatment (SD = 15.22) and 26.52 in the *Video* treatment (SD = 13.70)(see Figure 3a and Table 2). A Kruskal–Wallis test showed that the treatments are not statistically significant different from each other ($p = 0.217$). Likewise, pairwise comparisons between our treatments revealed no significant differences (see Table 2).¹⁴ Thus, more non-verbal cues during the virtual meeting did not facilitate higher team coordination in Round 1.

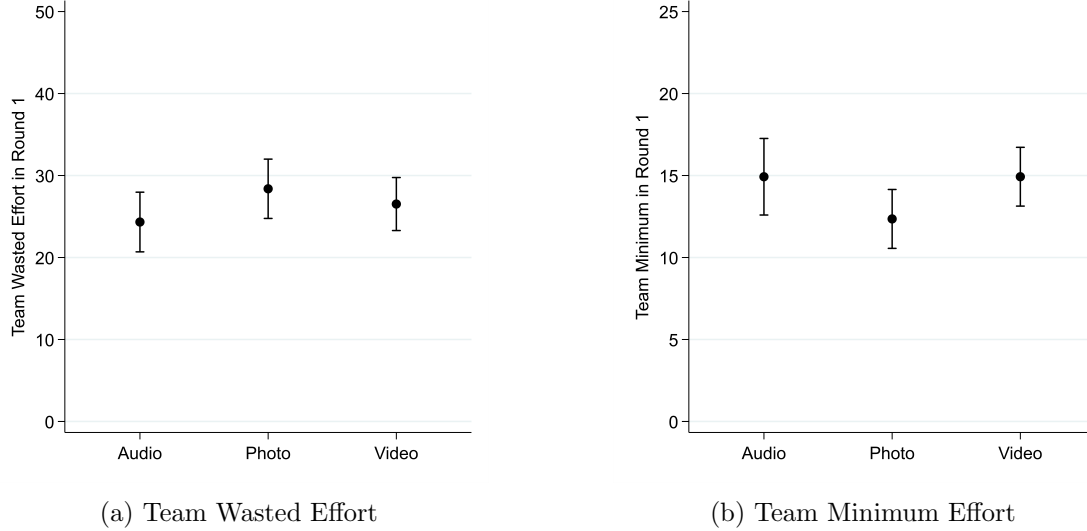


Figure 3: *Team Wasted Effort* (Coordination) and *Team Minimum* (Cooperation) in Round 1 by treatment

A series of linear regressions reported in Table 1 support the initial impression that enriching virtual communication with non-verbal cues did not decrease the wasted effort in teams. The coefficients for the *Photo* and *Video* treatment are positive in models (1) and (2), but the differences are not significant. Two-sided F-tests of the linear hypothesis $Photo = Video$ also showed that the difference between the *Photo* and the *Video* treatment is not statistically significant (Model (1) $p = 0.453$; Model (2) $p = 0.451$).

Result 1 Enriching virtual pre-play communication to transmit more non-verbal cues did not lead to higher levels of coordination (lower levels of wasted effort) in teams in the first round of a weakest-link game.

4.2 Cooperation

Descriptive information about the teams' cooperative behavior in Round 1 reveals mixed results. Recall that we measure cooperation with *Team Minimum Effort*. Thus, higher levels of cooperation are measured by higher levels of *Team Minimum Effort*. *Team Minimum Effort* was, on average, 14.93 in the *Audio* treatment

¹⁴Only the pairwise comparison between the *Audio* and *Photo* treatment showed a significant difference at the 10% level ($p = 0.095$).

Dep. Var.:	<i>Team Wasted Effort</i>		<i>Team Minimum Effort</i>	
	(1)	(2)	(3)	(4)
Photo Treatment	4.054 (2.618)	4.252 (2.666)	-2.572* (1.503)	-2.734* (1.505)
Video Treatment	2.193 (2.484)	2.350 (2.535)	0.002 (1.502)	-0.123 (1.522)
Homophily		7.829 (9.635)		3.101 (5.160)
Team Cohesion Part 1		-0.467 (1.391)		0.217 (0.795)
Team Risk Aversion		0.289 (0.793)		-0.513 (0.444)
Constant	24.328*** (1.857)	18.922* (10.021)	14.925*** (1.191)	15.492*** (5.573)
Observations	204	204	204	204
<i>Photo Treatment = Video Treatment</i>	0.453	0.451	0.048	0.046

Robust standard errors clustered at the team level in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
As pre-registered, we also conducted ordered probit regressions. The regression results are similar and can be found in Table 6 in Appendix A.

Table 1: Linear regression with *Team Wasted Effort* and *Team Minimum Effort* in Round 1 as dependent variables

	(1) Audio	(2) Photo	(3) Video	(1) vs (2)	(1) vs (3)	(2) vs (3)
Team Wasted Effort Round 1	24.33 (15.20)	28.38 (15.22)	26.52 (13.70)	$p = 0.095$	$p = 0.251$	$p = 0.487$
Team Minimum Effort Round 1	14.93 (9.75)	12.35 (7.55)	14.93 (7.60)	$p = 0.197$	$p = 0.596$	$p = 0.046$

Standard deviations are in parentheses. The last three columns report results from pairwise statistical comparison based on two-sided Mann-Whitney U tests.

Table 2: Summary statistics and pairwise treatment comparisons based on non-parametric tests

(SD = 9.75), 12.35 in the *Photo* treatment (SD = 7.55) and 14.93 in the *Video* treatment (SD = 7.60). While a Kruskal–Wallis test showed that these treatment differences are not statistically significant ($p = 0.137$), the *Photo* treatment yielded the worst results (see Figure 3b and Table 2).¹⁵

A series of linear regressions reported in Table 1 support the initial impression of an U-shaped form of how non-verbal cues affected team cooperation. The coefficient for the *Photo* treatment is negative in Models (3) and (4) and significant at the 10% level. In our naturalistic “weakest-link” game, adding subjects’ selfies in an audio meeting led to a reduction of the minimum hours allocated to the team project by an average of more than 2.5 hours per week (approx. -17%). In contrast, the coefficient for the *Video* treatment is close to zero in Models (3) and (4) and not significant. Two-sided F-tests of the linear hypothesis *Photo* = *Video* showed that the difference between the *Photo* and the *Video* treatment is statistically significant (Model (3) $p = 0.048$; Model (4) $p = 0.046$). Overall, the *Team Minimum Effort* was higher in both the *Audio* and the

¹⁵The pairwise comparison between the *Photo* and *Video* treatment showed a significant difference at the 5% level ($p = 0.046$).

Video treatment compared to the *Photo* treatment. Thus, the *Audio* and the *Video* treatment led to similar results. While static non-verbal cues decreased cooperation, allowing dynamic non-verbal cues appears to have mitigated the backfiring compared to the *Audio* treatment.

Result 2 Enriching virtual pre-play communication to transmit more non-verbal cues revealed an U-shaped impact on cooperation in teams in the first round of a weakest-link game. Adding subjects' selfies as static non-verbal cues had a negative impact, whereas dynamic non-verbal cues led to similar results as audio communication.

4.3 Exploratory Analysis

In Section 4.1 and 4.2, we analyzed decisions on team-level in Round 1 of the weakest-link game. In this section, we provide exploratory analyses to assess decisions on subject-level, decisions over all 10 rounds on team-level and potential mechanism for the results in Section 4.1 and 4.2.

To investigate cooperation and coordination behavior on subject-level, we followed the literature (e.g., Avoyan and Ramos, 2023) and calculated the normalized efficiency of subjects for each round in the weakest-link game.

Efficiency in round r

$$\text{efficiency (subject-level)} = \frac{\text{actual payoff} - \text{minimum possible payoff}}{\text{maximum possible payoff} - \text{minimum possible payoff}} = \frac{\text{actual payoff}}{400}$$

In our study, the minimum possible payoff in the weakest-link games was 0 ECU and the maximum possible payoff was 400 ECU. In Round 1, *Efficiency* was, on average, 0.585 in the *Audio* treatment, 0.536 in the *Photo* treatment, and 0.576 in the *Video* treatment (see Figure 6 in Appendix A). Similar to the results on team cooperation in Round 1, *Efficiency* was statistically significantly higher in *Audio* and *Video* compared to *Photo* (see Table 4).

Again, we conducted a series of linear regressions reported in Table 3 to investigate the robustness of this initial impression. The coefficient for *Photo* is negative and statistically significant at the 10% level, also when controlling for demographic characteristics as well as individual risk and group preferences.¹⁶ Two-sided F-tests of the linear hypothesis $Photo = Video$ showed similar results, but differences are no longer significant at the 10% level when including control variables (Model (1) $p = 0.090$; Model (2) $p = 0.106$). Thus, similar to the results on cooperation at the team-level, the performance at the subject-level, measured in terms of *Efficiency*, was reduced by including static cues compared to no non-verbal cues. Dynamic cues in *Video* enhanced subjects' performance again to the level of *Audio*.

¹⁶Remember that the pairwise comparison between the *Photo* and *Video* treatment revealed a significant difference in *Team Cohesion Part 1* at the 10% level (Mann-Whitney U test, $p = 0.069$).

Dep. Var.: <i>Efficiency</i>	(1)	(2)
Photo Treatment	-0.049*	-0.049*
	(0.027)	(0.027)
Video Treatment	-0.009	-0.010
	(0.026)	(0.026)
Female		0.016
		(0.015)
Age		-0.000
		(0.001)
Ethnicity White		0.019
		(0.017)
Team Cohesion Part 1		-0.003
		(0.006)
Risk Aversion		-0.002
		(0.003)
Constant	0.585***	0.592***
	(0.021)	(0.048)
Observations	612	612
<i>Photo Treatment = Video Treatment</i>	0.090	0.106

Robust standard errors clustered at the team level in parentheses,
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Linear regression on subject-level with *Efficiency* in Round 1 as dependent variable

Following this subject-level analysis, we analyzed the team behavior over all ten rounds of the weakest-link game. For *Team Wasted Effort* over all rounds, no distinct differences between the treatments were observed (see Figure 4a). While *Team Wasted Effort* in the *Photo* treatment stayed highest until Round 3 and the order of the other two treatments also remained the same in the first three rounds, the treatment differences declined and, starting with Round 4, no clear order or treatment differences were visible. Over the rounds, the average *Team Wasted Effort* in all treatments decreased significantly (see Table 7).

Figure 4b shows that the treatment differences in *Team Minimum Effort* were largest in the first round and disappeared in round three of the weakest-link game. From then on, there was no longer a distinct difference between treatments. In Round 2, the treatment difference described in Section 4.2 between *Photo* and *Video* was no longer statistically significant (two-sided Mann-Whitney U test, $p = 0.297$). The average *Team Minimum Effort* across all rounds was 14 in *Audio*, 13.54 in *Photo*, and 13.93 in *Video* and all pairwise comparisons based on two-sided Mann-Whitney U tests are not statistically significant (see Table 4). Overall, cooperative behavior is stable and the *Team Minimum Effort* does not increase or decrease significantly over the rounds.

Next, we discuss potential mechanisms that may explain our findings on team cooperation in Round 1. The results from Gächter et al. (2024) showed that beliefs and team cohesion affect team cooperation in a weakest-link game. Thus, as described in Section 2.3.3, we elicited subjects' beliefs about the effort levels of their other two team members after the first round and team cohesion in the post-experimental survey. To

	(1) Audio	(2) Photo	(3) Video	(1) vs (2)	(1) vs (3)	(2) vs (3)
Individual Efficiency Round 1	0.585 (0.196)	0.536 (0.184)	0.576 (0.171)	$p = 0.032$	$p = 0.840$	$p = 0.037$
Team Minimum Effort Round 1-10	14 (9.63)	13.54 (8.86)	13.93 (9.27)	$p = 0.952$	$p = 0.954$	$p = 0.956$
Team Belief Minimum Round 1	19.75 (6.87)	19.36 (6.820)	20.48 (6.65)	$p = 0.946$	$p = 0.395$	$p = 0.385$
Team Cohesion Part 2	3.92 (0.97)	3.59 (1.01)	3.88 (1.03)	$p = 0.054$	$p = 0.951$	$p = 0.064$
Team Anonymity	5.68 (0.69)	5.41 (0.74)	5.17 (0.84)	$p = 0.038$	$p = 0.000$	$p = 0.054$
Team Human Presence	4.74 (0.92)	4.47 (0.90)	4.74 (0.80)	$p = 0.066$	$p = 0.851$	$p = 0.070$

Standard deviations are in parentheses. The last three columns report results from pairwise statistical comparison based on two-sided Mann-Whitney U tests.

Table 4: Summary statistics for exploratory analyses and pairwise treatment comparisons based on non-parametric tests

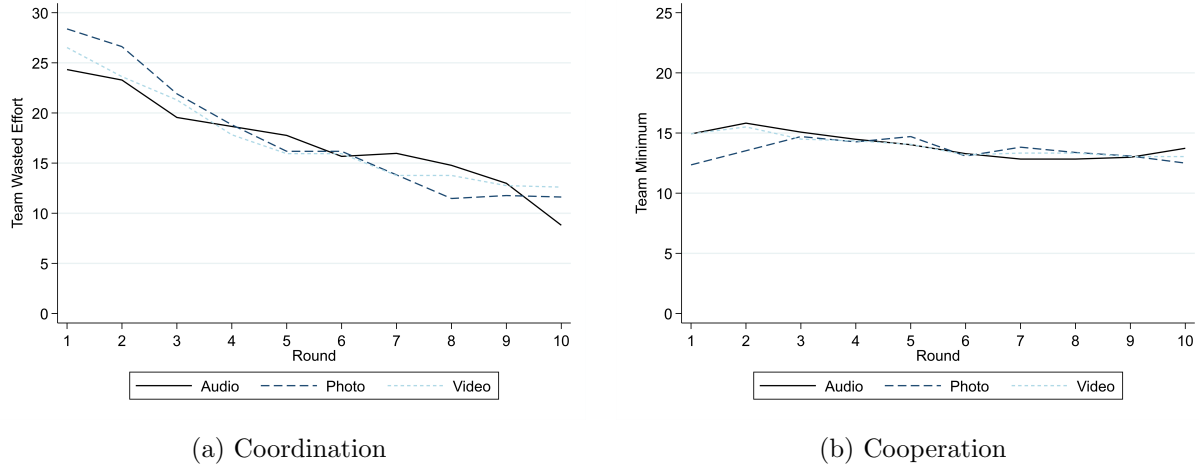


Figure 4: *Team Wasted Effort* and *Team Minimum* over all rounds by treatment

study beliefs and team cohesion as potential mechanisms on team-level, we followed (Gächter et al., 2024) and calculated a belief regarding the team minimum (*Team Belief Minimum Round 1*) and constructed a team cohesion index (*Team Cohesion Part 2*). In addition, we elicited subjects’ perceived anonymity (Hite et al., 2014) and their perceived human presence (Cohn et al., 2022), because our treatments vary subjects’ visual identification within the communication stage. *Team Anonymity* describes the team average of the subjects’ perceived anonymity, and *Team Human Presence* describes the team average of subjects’ perceived human presence. Values are higher if subjects perceived a higher anonymity or higher human presence.

Note that these four measures were not exogenous because they were influenced by our treatment variations. While *Team Belief Minimum Round 1* did not differ significantly between treatments (Kruskal-Wallis test, $p = 0.606$) and also pairwise comparisons are not significant (see Table 4), team cohesion varied. *Team Cohesion Part 2* was, on average, 3.92 in the baseline, 3.59 in the *Photo* treatment, and 3.88 in the *Video*

treatment.¹⁷ A Kruskal–Wallis test shows that these treatment differences were significant at the 10 % level ($p = 0.092$). Pairwise comparisons yield similar differences between the treatments in *Team Minimum* in the first round and show that static photos tended to mute the development of team cohesion (see Table 4 and Figure 7). Although a Kruskal–Wallis test reveals no significant difference ($p = 0.107$), the treatment differences for perceived human presence show similar patterns.

Data on *Team Anonymity* reveal that our treatment variation on visual anonymity had the expected effect on perceived anonymity. *Team Anonymity* was, on average, 5.68 in the baseline, 5.41 in the *Photo* treatment, and 5.17 in the *Video* treatment. Thus, the perceived anonymity of the subjects decreased significantly when more non-verbal cues for visual identification of other team members were available during the communication phase (Jonckheere–Terpstra test, $p < 0.001$).

Despite these results showing how non-verbal cues in virtual communication can influence team cohesion, human presence, and anonymity, we did not include these variables in the regressions in Table 1 due to endogeneity concerns. Hence, to investigate potential mechanisms of how our treatments affected the *Team Minimum Effort*, we conducted a decomposition analysis for each treatment. An analysis of correlation independent of our treatments shows that there is a positive significant correlation between *Team Minimum Effort* and *Team Belief Minimum Round 1* ($r = 0.6482$) and between *Team Minimum Effort* and *Team Cohesion Part 2* ($r = 0.2144$) (see Table 8). Therefore, the decomposition analysis in each treatment was based on the following simultaneous equation model where i indexes teams:

$$\begin{aligned} \text{Team Minimum Effort}_i = & \beta_0 + \beta_1 \cdot \text{Team Belief Minimum Round 1}_i + \beta_2 \cdot \text{Team Cohesion Part 2}_i \\ & + \beta_3 \cdot \text{Team Anonymity}_i + \beta_4 \cdot \text{Team Human Presence}_i \\ & + \beta_5 \cdot \text{Homophily}_i + \beta_6 \cdot \text{Team Risk Aversion}_i + \epsilon_i \end{aligned} \quad (1)$$

$$\begin{aligned} \text{Team Belief Minimum Round 1}_i = & \beta_0 + \beta_1 \cdot \text{Team Cohesion Part 2}_i + \beta_2 \cdot \text{Team Anonymity}_i \\ & + \beta_3 \cdot \text{Team Human Presence}_i + \beta_4 \cdot \text{Homophily}_i \\ & + \beta_5 \cdot \text{Team Risk Aversion}_i + \epsilon_i \end{aligned} \quad (2)$$

$$\begin{aligned} \text{Team Cohesion Part 2}_i = & \beta_0 + \beta_1 \cdot \text{Team Anonymity}_i + \beta_2 \cdot \text{Team Human Presence}_i \\ & + \beta_3 \cdot \text{Homophily}_i + \beta_4 \cdot \text{Team Risk Aversion}_i + \epsilon_i \end{aligned} \quad (3)$$

The results in Table 5 reveal that in all treatments only the *Team Belief Minimum Round 1* had a positive

¹⁷Interestingly, the average *Team Cohesion Part 2* was $\in (2, 4]$ in all three treatments. Thus, although the subjects did not know any other team members and the teams varied with respect to demographic characteristics, they achieved a medium level of team cohesion based on the categorization following Gächter et al. (2024).

	(1) <i>Audio</i>	(2) <i>Photo</i>	(3) <i>Video</i>
Equation 1			
Dep. Var.: Team Minimum Effort			
Team Belief Minimum Round 1	1.083*** (0.140)	0.637*** (0.106)	0.726*** (0.104)
Team Cohesion Part 2	-0.779 (1.135)	-0.653 (1.007)	0.577 (0.966)
Team Anonymity	-1.324 (1.286)	-1.424 (0.949)	0.428 (0.830)
Team Human Presence	0.784 (1.108)	1.138 (1.141)	0.138 (1.214)
Homophily	5.229 (6.505)	-7.719 (8.075)	4.558 (7.087)
Team Risk Aversion	-0.788 (0.644)	-0.002 (0.554)	0.065 (0.550)
Constant	2.322 (10.902)	10.231 (9.350)	-8.547 (8.939)
Equation 2			
Dep. Var.: Team Belief Minimum Round 1			
Team Cohesion Part 2	3.213*** (1.199)	2.432*** (0.939)	1.623 (1.220)
Team Anonymity	0.756 (1.198)	0.336 (1.170)	-0.151 (1.069)
Team Human Presence	-1.175 (1.111)	-1.294 (0.981)	-0.322 (1.517)
Homophily	1.054 (5.609)	18.972*** (6.388)	-10.088 (7.479)
Team Risk Aversion	-0.332 (0.626)	-0.315 (0.647)	-0.434 (0.469)
Constant	10.051 (8.982)	3.872 (8.175)	26.162*** (8.858)
Equation 3			
Dep. Var.: Team Cohesion Part 2			
Team Anonymity	0.018 (0.141)	0.005 (0.126)	0.074 (0.101)
Team Human Presence	0.751*** (0.097)	0.689*** (0.111)	0.986*** (0.101)
Homophily	-0.215 (0.592)	-0.568 (0.878)	-0.078 (0.705)
Team Risk Aversion	-0.047 (0.051)	0.134 (0.090)	0.072 (0.052)
Constant	0.723 (0.979)	-0.041 (1.141)	-1.602* (0.921)
Log pseudolikelihood	-734.5	-735.0	-738.1
Observations	67	68	69

Robust standard errors clustered at the team level in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Mediation analysis: Estimating simultaneous linear equations

significant effect on the team minimum in Round 1, while all other control variables, including *Team Cohesion Part 2*, had no significant direct effect on team cooperation in the first round (equation (1)). However, *Team Cohesion Part 2* indirectly affected team cooperation via beliefs but only in the *Audio* and *Photo* treatment. In Model (1) and (2), it had a positive significant effect on *Team Belief Minimum Round 1*, whereas in the *Video* treatment, team cohesion did not significantly affect teams' beliefs (equation (2)). When looking at the effect of *Homophily* on beliefs, we do see differences between the treatments. While in the *Audio* and *Video* treatments, *Homophily* had no significant effect on the minimum belief within teams, it had a positive and strongly significant effect in the *Photo* treatment. Compared to the baseline, static selfies made demographics more salient in a team, which might have increased the relevance of homophily on behavior within a team. In the *Video* treatment compared to the *Photo* treatment demographics of team members should be equally salient in a team. Still, the coefficient of *Homophily* in the *Video* treatment is not significant and, interestingly, negative and not positive compared to the *Photo* treatment. Thus, the dynamic non-verbal cues influenced the effect of *Homophily* on *Team Belief Minimum Round 1*. The third equation in Table 5 shows that in all three treatments only *Team Human Presence* had a positive and significant effect on team cohesion. It, therefore, indirectly affects team cooperation in the *Audio* and *Photo* treatment.

5 Discussion

The results of our study show that non-verbal cues affect team cooperation but not coordination. Against our expectation that an increased amount of non-verbal cues in general, as well as the possibility of identification alone, would lead to higher levels of coordination and cooperation, we even observed negative effects of visual identification through static non-verbal cues in the form of a selfie on team cooperation. In the following, we first discuss the robustness and possible mechanisms for our results on cooperation. Second, we explain limitations and highlight avenues for future research.

Previous research showed that high perceived attractiveness of the person displayed can affect the behavior of the other team members (Andreoni and Petrie, 2008; Ruffle and Shtudiner, 2015; Wilson and Eckel, 2006), e.g., through an increase in perceived cooperativeness (Zylbersztejn et al., 2024a). While we have subjective ratings of team members' physical attractiveness in the *Photo* and *Video* treatment, the usage of these scores comes with endogeneity problems. Therefore, we have conducted a short follow-up study to elicit exogenous ratings of subjects' perceived attractiveness (see Appendix B). *Team Attractiveness* describes the team average of exogenous ratings on subjects' attractiveness, *Team Variance Attractiveness* the variance of team members' attractiveness scores.¹⁸ The exogenous attractiveness ratings of our follow-up study and the

¹⁸The discussion of *Team Variance Attractiveness* can be found in Appendix B.

endogenous ratings in Part 2 are significantly correlated (Pearson correlations, $p < 0.01$). This holds true for individual attractiveness ratings, as well as for the *Team Attractiveness* and there are exogenous treatment differences for individual attractiveness ratings between *Photo* and *Video* (see Table 9 in Appendix A). We included *Team Attractiveness* as a control in the linear regressions (see Table 12 in Appendix A).¹⁹ It has no significant impact on the *Team Minimum Effort* in Round 1, and the coefficient for the *Video* treatment is positive in Model (5) and significant at the 10% level. Thus, *Team Attractiveness* cannot explain the differences between the *Photo* and *Video* treatment on cooperation. A subgroup analysis revealed that the significant treatment differences observed in the overall sample persisted only in teams with below-median *Team Attractiveness* (two-sided Mann-Whitney U test, $p = 0.043$). In contrast, no significant treatment differences were found in teams with above-median *Team Attractiveness* ($p = 0.412$). These findings were also confirmed with linear regressions (see Table 13). In teams that were below-median in terms of *Team Attractiveness*, treatment differences are significant at the 5% level. Including control variables, treatment differences remain, but only at the 10% level. In addition, *Homophily* had a direct significantly positive effect on team cooperation (at the 10% level). Therefore, the selfies particularly seem to have an effect in teams that are rated lower in terms of attractiveness, and in these teams increased similarity in terms of gender, age, and ethnicity was advantageous for cooperation.

In the past, it has been shown that the cooperation of others can be detected more accurately when seeing them in a video rather than a photo (Bonnefon et al., 2017). This could explain why the *Video* does better than the *Photo* treatment. Furthermore, some literature provides possible explanations for why cooperation in the *Photo* treatment was lower than in *Audio*. The presence of photos has been shown to be disadvantageous when inferring cooperation behavior in a trust game (Efferson and Vogt, 2013). Another possible explanation for why identification through selfies led to significantly lower cooperation is a potential distraction from verbal communication. Through the static nature of the photos, the communication stage could be altered, e.g., through a change in communication content or dynamics. This could be related to the lower perceived human presence in the *Photo* treatment. We analyzed the verbal content using the collected audio recordings to investigate possible treatment differences in the communication stage (see Appendix C for details). First, we looked at *Character Count*, which describes the number of characters used in a team (communication length). Figure 8a shows that the *Character Count* increased from the *Audio* to the *Photo* and further to the *Video* treatment. In particular, the teams in the *Video* treatment talked more than in the other two treatments, and these differences are significant (see Table 15). Similar to *Team Cohesion Part 2*, *Team Human Presence*, and *Team Anonymity*, we could not include *Character Count* in the regressions in Table 1 due to endogeneity concerns. However, in Table 16 in Appendix C we observe that independent of our treatments, *Character Count* had a significant positive effect on *Team Minimum Effort* in Round 1

¹⁹We only included *Team Attractiveness* as control variable in the regressions as pairwise correlations between the constructs elicited in the follow-up study are significant (see Table 10).

at the 5% level. Since *Character Count* was significantly higher in the *Video* treatment, the communication length might explain some treatment differences in cooperation between the *Photo* and the *Video* treatment. In addition, we used the *Standard Deviation Speech Share* to examine whether unequal speaking times, e.g., when one team member spoke for the majority of the seven-minute virtual meeting, influenced the subsequent results. While the positive effect of *Character Count* is robust when controlling for *Standard Deviation Speech Share* and other variables, *Standard Deviation Speech Share* had no significant impact on the *Team Minimum Effort* in Round 1. Moreover, *Standard Deviation Speech Share* did not significantly differ between treatments (see Table 15 in Appendix C). Additionally, the communication content was compared between teams and treatments, using clustering and topic-modeling approaches (see Appendix C for details on the analyses). As no treatment differences in communication content could be observed, the communication content, could not help understand the treatment differences.

Next, we outline some limitations of our study and propose directions for future research. First, a key limitation of our study is that the photos used are simple selfies shot by the subjects themselves and not professional photos, which are usually used in these circumstances. Therefore, results may not be directly transferable to practical applications. Moreover, we did not elicit or account for subjects' preferences and prior experiences with the specific meeting formats. The subjects may have been more accustomed to audio-only meetings, as many employees in small team settings typically keep their cameras off during meetings (Tolliver and Sass, 2024). Consequently, the unfamiliarity of the meeting format could have impacted their decisions. Furthermore, removing the usual option to turn the camera on or off might have led to reactance effects (Brehm, 1966). These reactance effects may have been particularly pronounced in cases where subjects felt that their personal comfort was restricted. Therefore, understanding and accounting for individual preferences and past experiences with these meeting formats could influence the results and help explain some of the observed behavior. This potential limitation relates explicitly to the *Photo* treatment. The lower cooperation observed in this treatment might be partially explained by subjects' lack of experience with audio meetings enriched with selfies. Given that many individuals may be more familiar with purely audio-based meetings or those with live video, this unfamiliar format could have hindered their engagement or sense of comfort, thus impacting their cooperative behavior. This effect might decline with increasing familiarity. Since existing research showed that the choice of a communication channel can matter for individuals (Nieken and Walther, 2024), future research could examine whether individuals prefer different meeting formats and whether such preferences enhance team outcomes. Expanding on the previous limitation, our study did not consider the potential effects of repeated communication within the same meeting format. It is plausible that repeated exposure to a specific format could alter subjects' behavior over time. For example, with repeated communication, subjects might become more accustomed to the *Photo* treatment, reducing initial discomfort and potentially leading to more collaboration. Future research could expand on this by

incorporating repeated communication and decision stages to explore how ongoing interaction in a consistent meeting format influences cooperation and coordination. This could provide valuable insights into how non-verbal cues, such as mimics and gestures, play a role in long-term collaboration. Additionally, future studies could focus on analyzing the non-verbal cues in the video meeting. As we only have audio recordings available, our study can not investigate how and which non-verbal cues actually affected and potentially improved cooperation.

Another possibility for future research involves investigating the impact of adding photos to audio meetings when communication is game-related, and subjects are more focused on achieving specific goals. Thereby adding to the existing literature on communication with promises and commitment.

6 Conclusion

Our study introduces a controlled online experiment investigating the impact of non-verbal cues in virtual communication on cooperation and coordination. Specifically, we focused on comparing audio communication, audio communication enhanced with selfies, and video communication. Overall, we did not find a significant effect of non-verbal cues on coordination in the first round of a weakest-link game. However, cooperation was significantly higher when communication did not have any non-verbal cues (*Audio*) or was enriched with dynamic non-verbal cues (*Video*) than when static non-verbal cues allow visual identification (*Photo*). Similar results hold for the individual efficiency of subjects' outcomes. Looking at potential mechanisms, our results suggest that the belief on the team minimum had a significant effect on the team minimum, while perceived team cohesion indirectly affected the team minimum via beliefs only in the *Audio* and *Photo* treatment and not when cameras were turned on.

Our results suggest that team cooperation was dampened through static non-verbal cues even though they enabled visual identification and thereby significantly decreased anonymity. A possible explanation as to why static non-verbal cues did not have the expected effect on cooperation is the perceived human presence which was marginally significantly lower in the *Photo* treatment compared to the other treatments and indirectly affected team cooperation. Thus, to increase cooperation in newly formed teams in a one-shot setting, it was not necessary to engage team members to turn on their cameras, as audio-only worked similarly well. Instead, encouraging teams to talk more can be effective, as more verbal communication has a positive effect on cooperation. Our insights reveal some practical implications for practitioners and policy-makers. To increase team cooperation as well as coordination, it should not matter if new virtual teams get to know each other in audio or video meetings. Still, based on our results, both of these meeting formats should be preferred over audio meetings enriched with selfies. This is particularly relevant for organizations and companies with very diverse teams (in relation to age, gender, and ethnicity), as cooperation in such teams might suffer from

highlighting these differences with selfies. Further research could study repeated interaction settings and extend the research on the effect of photos when communication is game-related.

References

- ANDREONI, J. AND R. PETRIE (2004): “Public goods experiments without confidentiality: a glimpse into fund-raising,” *Journal of Public Economics*, 88, 1605–1623.
- (2008): “Beauty, gender and stereotypes: Evidence from laboratory experiments,” *Journal of Economic Psychology*, 29, 73–93.
- ARECHAR, A. A., S. GÄCHTER, AND L. MOLLEMAN (2018): “Conducting interactive experiments online,” *Experimental Economics*, 21, 99–131.
- ARON, A., E. N. ARON, AND D. SMOLLAN (1992): “Inclusion of Other in the Self Scale and the structure of interpersonal closeness,” *Journal of Personality and Social Psychology*, 63, 596–612.
- ASH, E. AND S. HANSEN (2023): “Text Algorithms in Economics,” *Annual Review of Economics*, 15, 659–688.
- AVOYAN, A. AND J. RAMOS (2023): “A Road to Efficiency through Communication and Commitment,” *American Economic Review*, 113, 2355–2381.
- BABUTSIDZE, Z., N. HANAKI, AND A. ZYLBERSZTEJN (2021): “Nonverbal content and trust: An experiment on digital communication,” *Economic Inquiry*, 59, 1517–1532.
- BALLIET, D. (2010): “Communication and cooperation in social dilemmas: A meta-analytic review,” *Journal of Conflict Resolution*, 54, 39–57.
- BARRERO, J. M., N. BLOOM, AND S. J. DAVIS (2023): “The Evolution of Work from Home,” *Journal of Economic Perspectives*, 37, 23–50.
- BLOOM, N., R. HAN, AND J. LIANG (2022): “How hybrid working from home works out,” Tech. rep., National Bureau of economic research.
- BOCHET, O., T. PAGE, AND L. PUTTERMAN (2006): “Communication and punishment in voluntary contribution experiments,” *Journal of Economic Behavior & Organization*, 60, 11–26.
- BOHNET, I. AND B. S. FREY (1999): “The sound of silence in prisoner’s dilemma and dictator games,” *Journal of Economic Behavior & Organization*, 38, 43–57.
- 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, 276–281.
- BRANDTS, J. AND D. J. COOPER (2006a): “A Change Would Do You Good An Experimental Study on How to Overcome Coordination Failure in Organizations,” *American Economic Review*, 96, 669–693.
- (2006b): “Observability and overcoming coordination failure in organizations: An experimental study,” *Experimental Economics*, 9, 407–423.
- BREDIN, H. AND A. LAURENT (2021): “End-To-End Speaker Segmentation for Overlap-Aware Resegmentation,” in *Proceedings Interspeech 2021*, 3111–3115.
- BREHM, J. W. (1966): *A theory of psychological reactance*, Oxford, England: Academic Press.
- BROSIG, J. AND J. WEIMANN (2003): “The Effect of Communication Media on Cooperation,” *German Economic Review*, 4, 217–241.
- CHARNESS, G. AND M. DUFWENBERG (2006): “Promises and partnership,” *Econometrica*, 74, 1579–1601.
- 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, 29–35.

- CHEN, D. L., M. SCHONGER, AND C. WICKENS (2016): “oTree—An open-source platform for laboratory, online, and field experiments,” *Journal of Behavioral and Experimental Finance*, 9, 88–97.
- CHEN, R. AND Y. CHEN (2011): “The Potential of Social Identity for Equilibrium Selection,” *American Economic Review*, 101, 2562–2589.
- CIALDINI, R. B., S. L. BROWN, B. P. LEWIS, C. LUCE, AND S. L. NEUBERG (1997): “Reinterpreting the empathy–altruism relationship: When one into one equals oneness,” *Journal of Personality and Social Psychology*, 73, 481–494.
- COHN, A., T. GESCHE, AND M. A. MARÉCHAL (2022): “Honesty in the digital age,” *Management Science*, 68, 827–845.
- COOPER, R., D. V. DEJONG, R. FORSYTHE, AND T. W. ROSS (1992): “Communication in coordination games,” *The Quarterly Journal of Economics*, 107, 739–771.
- CRAWFORD, V. (1998): “A survey of experiments on communication via cheap talk,” *Journal of Economic theory*, 78, 286–298.
- DAFT, R. L. AND R. H. LENGEL (1986): “Organizational information requirements, media richness and structural design,” *Management Science*, 32, 554–571.
- DONG, L., M. MONTERO, AND A. POSSAJENNIKOV (2018): “Communication, leadership and coordination failure,” *Theory and Decision*, 84, 557–584.
- DRISKELL, J. E., E. SALAS, AND S. HUGHES (2010): “Collective Orientation and Team Performance: Development of an Individual Differences Measure,” *Human Factors*, 52, 316–328.
- ECKEL, C. C. AND R. PETRIE (2011): “Face value,” *American Economic Review*, 101, 1497–1513.
- EFFERSON, C. AND S. VOGT (2013): “Viewing men’s faces does not lead to accurate predictions of trustworthiness,” *Scientific Reports*, 3, 1047.
- FALK, A., A. BECKER, T. DOHMEN, D. HUFFMAN, AND U. SUNDE (2023): “The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences,” *Management Science*, 69, 1935–1950.
- FAUVILLE, G., M. LUO, A. C. M. QUEIROZ, A. LEE, J. N. BAIENSON, AND J. HANCOCK (2023): “Video-conferencing usage dynamics and nonverbal mechanisms exacerbate Zoom Fatigue, particularly for women,” *Computers in Human Behavior Reports*, 10, 100271.
- FROHLICH, N. AND J. OPPENHEIMER (1998): “Some consequences of e-mail vs. face-to-face communication in experiment,” *Journal of Economic Behavior & Organization*, 35, 389–403.
- GENTZKOW, M., B. KELLY, AND M. TADDY (2019): “Text as Data,” *Journal of Economic Literature*, 57, 535–74.
- GRÖZINGER, N., B. IRLENBUSCH, K. LASKE, AND M. SCHRÖDER (2020): “Innovation and communication media in virtual teams—An experimental study,” *Journal of Economic Behavior & Organization*, 180, 201–218.
- GÄCHTER, S., C. STARMER, AND F. TUFANO (2024): “Measuring “Group Cohesion” to Reveal the Power of Social Relationships in Team Production,” *The Review of Economics and Statistics*, 1–16.
- HAGEMANN, V. (2020): “Deutschsprachiges Instrument zur Erfassung der Kollektiven Orientierung (KO),” *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*.
- HE, S., T. OFFERMAN, AND J. VAN DE VEN (2017): “The sources of the communication gap,” *Management Science*, 63, 2832–2846.

- HITE, D. M., T. VOELKER, AND A. ROBERTSON (2014): “Measuring Perceived Anonymity: The Development of a Context Independent Instrument,” *Journal of Methods and Measurement in the Social Sciences*, 5, 22–39.
- KAISER, N., K. HENRY, AND H. EYJÓLFSDÓTTIR (2022): “Eye Contact in Video Communication: Experiences of Co-creating Relationships,” *Frontiers in Psychology*, 13.
- KOUDENBURG, N., T. POSTMES, AND E. H. GORDIJN (2013): “Conversational Flow Promotes Solidarity,” *PLOS ONE*, 8, e78363.
- LEINER, D. J. (2024): “SoSci Survey (Version 3.4.00) [Computer software],” <https://www.soscisurvey.de>.
- LEWIS, M., Y. LIU, N. GOYAL, M. GHAZVININEJAD, A. MOHAMED, O. LEVY, V. STOYANOV, AND L. ZETTLEMOYER (2019): “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension,” *arXiv: 1910.13461*.
- MANZINI, P., A. SADRIEH, AND N. J. VRIEND (2009): “On smiles, winks and handshakes as coordination devices,” *The Economic Journal*, 119, 826–854.
- NIEKEN, P. (2023): “Charisma in the gig economy: The impact of digital leadership and communication channels on performance,” *The Leadership Quarterly*, 34, 101631.
- NIEKEN, P. AND S. WALTHER (2024): “Honesty in Virtual Communication,” .
- NIESSEN, D., C. BEIERLEIN, B. RAMMSTEDT, AND C. M. LECHNER (2020): “An English-language adaptation of the Interpersonal Trust Short Scale (KUSIV3),” *Measurement Instruments for the Social Sciences*, 2, 10.
- QUEIROZ, A. C. M., A. Y. LEE, M. LUO, G. FAUVILLE, J. T. HANCOCK, AND J. N. BAIENSON (2023): “Too tired to connect: Understanding the associations between video-conferencing, social connection and well-being through the lens of zoom fatigue,” *Computers in Human Behavior*, 149, 107968.
- RADFORD, A., J. W. KIM, T. XU, G. BROCKMAN, C. MCLEAVEY, AND I. SUTSKEVER (2023): “Robust Speech Recognition via Large-Scale Weak Supervision,” in *Proceedings of the 40th International Conference on Machine Learning*, vol. 202 of *Proceedings of Machine Learning Research*, 28492–28518.
- ŘEHŮŘEK, R. AND P. SOJKA (2010): “Software Framework for Topic Modelling with Large Corpora,” in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, ELRA, 45–50.
- RIEDL, A., I. M. T. ROHDE, AND M. STROBEL (2016): “Efficient Coordination in Weakest-Link Games,” *The Review of Economic Studies*, 83, 737–767.
- RÖDER, M., A. BOTH, AND A. HINNEBURG (2015): “Exploring the Space of Topic Coherence Measures,” in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, Association for Computing Machinery, WSDM ’15, 399–408.
- ROUSSEUW, P. J. (1987): “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *Journal of Computational and Applied Mathematics*, 20, 53–65.
- RUFFLE, B. J. AND Z. SHTUDINER (2015): “Are Good-Looking People More Employable?” *Management Science*, 61, 1760–1776.
- SPRECHER, S. (2014): “Initial interactions online-text, online-audio, online-video, or face-to-face: Effects of modality on liking, closeness, and other interpersonal outcomes,” *Computers in Human Behavior*, 31, 190–197.
- TODOROV, A., A. N. MANDISODZA, A. GOREN, AND C. C. HALL (2005): “Inferences of Competence from Faces Predict Election Outcomes,” *Science*, 308, 1623–1626.
- TOKU, H., T. SHICHIJO, AND K. OGAWA (2022): “The effect of continuous-time cheap talk in the experimental minimum effort game,” *Applied Economics Letters*, 29, 687–699.

- TOLLIVER, M. AND J. SASS (2024): “Hybrid Work Has Changed Meetings Forever,” *Harvard Business Review*.
- VAN HUYCK, J. B., R. C. BATTALIO, AND R. O. BEIL (1990): “Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure,” *The American Economic Review*, 80, 234–248.
- VAN LEEUWEN, B., C. N. NOUSSAIR, T. OFFERMAN, S. SUETENS, M. VAN VEELLEN, AND J. VAN DE VEN (2018): “Predictably angry—Facial cues provide a credible signal of destructive behavior,” *Management Science*, 64, 3352–3364.
- VYOPTA (2022): “The Challenges of Hybrid Work: Vyopta 2022 Survey,” https://go.vyopta.com/vyopta_wakefield_survey_22.
- WERNER, V. AND M. ANDRES (2024): “Destructive Communication,” *Available at SSRN: 4836117*.
- WILSON, R. K. AND C. C. ECKEL (2006): “Judging a book by its cover: Beauty and expectations in the trust game,” *Political Research Quarterly*, 59, 189–202.
- YIN, W., J. HAY, AND D. ROTH (2019): “Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach,” *arXiv: 1909.00161*.
- 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.
- (2021): “Predicting trustworthiness across cultures: An experiment,” *Frontiers in Psychology*, 12, 727550.
- ZYLBERSZTEJN, A., Z. BABUTSIDZE, N. HANAKI, AND A. HOPFENSITZ (2024a): “How beautiful people see the world: Cooperativeness judgments of and by beautiful people,” *Journal of Economic Behavior & Organization*, 218, 296–308.
- ZYLBERSZTEJN, A., Z. BABUTSIDZE, N. HANAKI, AND M.-S. ROUL (2024b): “Anonymity, nonverbal communication and prosociality in digitized interactions: An experiment on charitable giving,” *Journal of Economic Psychology*, 105, 102769.

A Tables and Figures

Please indicate how much you would enjoy the following types of vacation ...

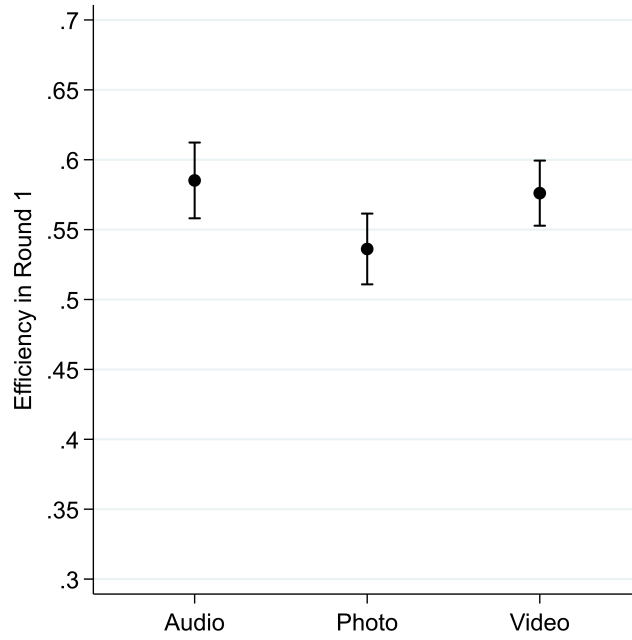
	not at all						totally
	1	2	3	4	5	6	7
Educational vacation.	☐	☐	☐	☐	☐	☐	☐
Cruise vacation.	☐	☐	☐	☐	☐	☐	☐
City vacation.	☐	☐	☐	☐	☐	☐	☐
Party vacation.	☐	☐	☐	☐	☐	☐	☐
Camping vacation.	☐	☐	☐	☐	☐	☐	☐
Winter sports vacation.	☐	☐	☐	☐	☐	☐	☐
Sun, sea and beach vacation.	☐	☐	☐	☐	☐	☐	☐

Figure 5: Rating of different vacations

Dep. Var.:	<i>Team Wasted Effort</i> (1)	<i>Team Wasted Effort</i> (2)	<i>Team Minimum Effort</i> (3)	<i>Team Minimum Effort</i> (4)
Photo Treatment	0.291 (0.183)	0.304 (0.185)	-0.321* (0.188)	-0.344* (0.188)
Video Treatment	0.164 (0.175)	0.175 (0.179)	0.021 (0.190)	0.005 (0.192)
Homophily		0.507 (0.658)		0.377 (0.660)
Team Cohesion Part 1		-0.029 (0.096)		0.034 (0.103)
Team Risk Aversion		0.022 (0.054)		-0.065 (0.057)
Observations	204	204	204	204
<i>Photo Treatment = Video Treatment</i>	0.465	0.463	0.048	0.044

Robust standard errors clustered at the team level in parentheses, * $p < 0.10$, ** $p < 0.05$,
*** $p < 0.01$

Table 6: Ordered probit regressions with *Team Wasted Effort* and *Team Minimum Effort* in Round 1 as dependent variables



Note: Dots indicate averages and whiskers indicate 95% confidence intervals.

Figure 6: *Efficiency* in Round 1 by treatment

Dep. Var.: Team Wasted Effort	(1)	(2)
Photo Treatment	0.497 (1.569)	0.919 (1.557)
Video Treatment	0.227 (1.679)	0.461 (1.669)
Round	-1.655*** (0.133)	-1.655*** (0.133)
Homophily		7.846 (5.470)
Team Cohesion Part 1		-1.159 (0.849)
Team Risk Aversion		0.603 (0.493)
Constant	26.280*** (1.395)	21.562*** (6.228)
Observations	2040	2040

Robust standard errors clustered at the team level in parentheses,

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Linear panel regressions (random effects) with *Team Wasted Effort* as dependent variable

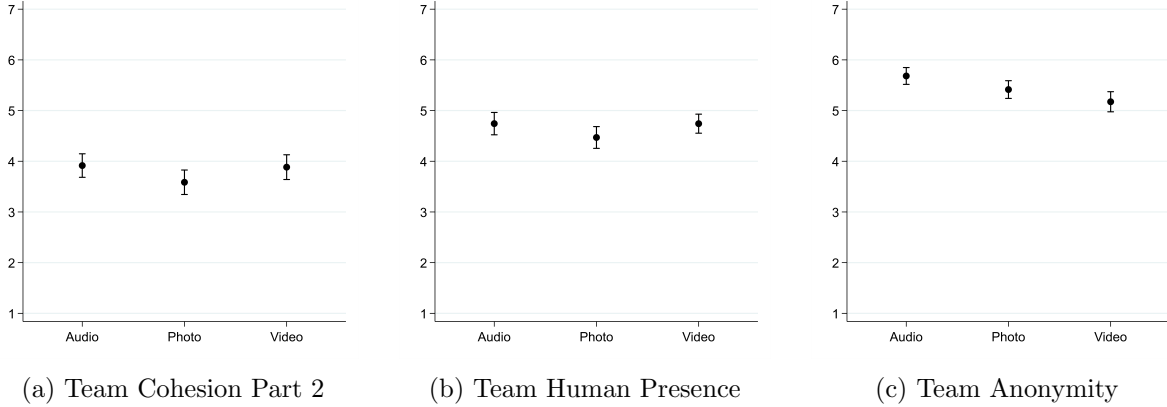


Figure 7: *Team Cohesion Part 2*, *Team Human Presence* and *Team Anonymity* by treatment

Variable	(1)	(2)	(3)	(4)
(1) Team Minimum Effort Round 1				
(2) Team Belief Minimum Round 1	0.648***			
(3) Team Cohesion Part 2	0.214***	0.267***		
(4) Team Anonymity	-0.059	-0.034	-0.083	
(5) Team Human Presence	0.155**	0.132*	0.689***	-0.186***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Pairwise correlations at the team-level ($n = 204$)

B Follow-up study to elicit exogenous ratings of subjects' selfies

In the follow-up study, we asked subjects who had not participated in any part of this study before to rate the selfies (taken in Part 1) of subjects that had taken part in the *Photo* and *Video* treatment. The raters were recruited from the same Prolific subject pool as our main study subjects, using the same screeners.²⁰ We randomly allocated the 411 selfies into sets of ten, while ensuring a 50-50 split between the men and women selfies (44 sets). To ensure an equal gender split, some selfies were used in two sets. Four random orders of the selfies were generated for each set (176 subsets) (Todorov et al., 2005); thus, the effects of the order of selfies within a set should be limited.²¹ To achieve a reliable rating database, we ensured that each of the 44 unique sets (of 10 selfies) was rated by a minimum of 24 different raters (12 men, 12 women), which resulted in a required sample of 1056 subjects.²²

Each rater was randomly assigned to one set and asked to rate the subjects displayed in the selfies according to their attractiveness (1: Attractive - 7: Not attractive), trustworthiness (1: Trustworthy - 7: Not trust-

²⁰With the exception of the agreement to the recording of videos.

²¹The only restriction was that an order should not start with all five selfies of the same gender.

²²Selfies included in two sets therefore had a minimum of 48 ratings and each subset was rated by at least three men and three women.

	(1) Photo	(2) Video	(1) vs (2)
Attractiveness	3.75 (0.61)	3.87 (0.60)	$p = 0.044$
Trustworthiness	4.09 (0.46)	4.19 (0.48)	$p = 0.033$
Intelligence	4.54 (0.52)	4.63 (0.50)	$p = 0.148$
Confidence	4.34 (0.67)	4.56 (0.64)	$p = 0.001$
Masculinity	4.19 (1.32)	4.06 (1.41)	$p = 0.267$
Competence	4.61 (0.48)	4.71 (0.46)	$p = 0.040$
Predictability	4.01 (0.37)	4.09 (0.37)	$p = 0.029$

Standard deviations are in parentheses. The last column reports results from pairwise statistical comparison based on two-sided Mann-Whitney U tests.

Table 9: Summary statistics of individual averages for exogenous selfie ratings and pairwise treatment comparisons based on non-parametric tests

worthy), intelligence (1: Not intelligent - 7: Intelligent), confidence (1: Not self-confident - 7: Self-confident), masculinity (1: Very feminine - 7: Very masculine), competence (1: Not competent - 7: Competent), and predictability (1: Predictable - 7: Not predictable) on a 7-point Likert scale, following Zylbersztejn et al. (2024a). After rating the ten selfies on all seven dimensions, we asked subjects whether they knew any of the persons they were shown and elicited demographic data. Subjects received a fixed payment of £2.00. Overall, 1062 raters (530 men, 528 women, 4 other) successfully completed the follow-up study,²³ resulting in 24 to 49 raters per selfie.

For the analysis, ratings for attractiveness, trustworthiness, and predictability were reverse-coded. For each *Photo* and *Video* treatment subject the mean attractiveness rating is calculated and used for subject-level analysis. For team-level analysis, the *Team Attractiveness* is the team average of the ratings on subjects' attractiveness. The same approach was used to compute *Team Trustworthiness*, *Team Intelligence*, *Team Confidence*, *Team Masculinity*, *Team Competence*, and *Team Predictability*. In addition, *Team Variance Attractiveness* described the variance of team members' attractiveness scores, to better capture the heterogeneity of team members. Similarly, we also calculated *Team Variance Trustworthiness*, *Team Variance Intelligence*, *Team Variance Confidence*, *Team Variance Masculinity*, *Team Variance Competence*, and *Team Variance Predictability*. While the analysis of *Team Attractiveness* is discussed in Section 5, *Team Variance Attractiveness* is investigated here.

We included the *Team Variance Attractiveness* as a control in the linear regression (see Table 12 in Appendix A).²⁴ The treatment differences between *Photo* and *Video* remained significant at the 5% level (see Model (6)). Therefore, *Team Variance Attractiveness* cannot explain the treatment differences. Still, *Team Variance Attractiveness* has a significant negative effect on cooperation in the *Photo* treatment at the 10% level. A subgroup analysis of teams with a *Team Variance Attractiveness* above or below the median revealed that

²³A total of 1099 subjects completed the study, but 37 subjects had to be excluded because they reported technical problems (30), knew at least one of the persons they were shown (4), or failed our attention check (3).

²⁴We only included *Team Variance Attractiveness* as control variable in the regressions as many pairwise correlations between the constructs elicited in the follow-up study are significant (see Table 11).

Variable	(1)	(2)	(3)	(4)	(5)	(6)
(1) Team Attractiveness						
(2) Team Trustworthiness	0.74***					
(3) Team Intelligence	0.62***	0.61***				
(4) Team Confidence	0.57***	0.48***	0.54***			
(5) Team Masculinity	-0.40***	-0.53***	-0.17**	-0.16*		
(6) Team Competence	0.64***	0.65***	0.90***	0.69***	-0.15*	
(7) Team Predictability	0.37***	0.59***	0.43***	0.19**	-0.37***	0.42***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Pairwise Pearson’s correlation coefficients at the team-level ($n = 137$) of exogenous selfie ratings: Team averages

Variable	(1)	(2)	(3)	(4)	(5)	(6)
(1) Team Variance Attractiveness						
(2) Team Variance Trustworthiness	0.31***					
(3) Team Variance Intelligence	0.16*	0.35***				
(4) Team Variance Confidence	0.40***	0.14*	0.16*			
(5) Team Variance Masculinity	0.15*	0.21**	-0.01	0.10		
(6) Team Variance Competence	0.17**	0.39***	0.81***	0.29***	-0.04	
(7) Team Variance Predictability	0.10	0.48***	0.31***	0.10	0.18**	0.30***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Pairwise Pearson’s correlation coefficients at the team-level ($n = 137$) of exogenous selfie ratings: Team variances

only for teams with a *Team Variance Attractiveness* above the median, treatment differences are significant (two-sided Mann-Whitney U test, $p = 0.047$). Linear regressions in these subgroups confirmed the results. Significant treatment differences are only observed in the subgroup of teams with an above-median variance in team members’ attractiveness scores (see Table 14).

Dep. Var. by Column:	<i>Team Wasted Effort</i>			<i>Team Minimum Effort</i>		
	(1)	(2)	(3)	(4)	(5)	(6)
Video Treatment	-1.861 (2.475)	-1.571 (2.533)	-2.078 (2.537)	2.575** (1.294)	2.517* (1.283)	2.655** (1.292)
Homophily		16.754 (11.328)	17.294 (11.662)		2.375 (6.398)	1.429 (6.318)
Team Cohesion Part 1		-1.733 (1.572)	-1.463 (1.580)		0.470 (0.842)	0.488 (0.822)
Team Risk Aversion		0.391 (0.873)	0.197 (0.871)		-0.303 (0.477)	-0.211 (0.475)
Team Attractiveness		-4.867 (3.308)			1.411 (1.749)	
Team Variance Attractiveness			1.388 (4.483)			-3.773* (2.239)
Constant	28.382*** (1.846)	40.236** (17.434)	21.249* (12.304)	12.353*** (0.916)	5.429 (9.035)	12.035* (6.307)
Observations	137	137	137	137	137	137

Robust standard errors clustered at the team level in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Linear regression with *Team Wasted Effort* and *Team Minimum Effort* in Round 1 as dependent variables, the *Team Attractiveness* or *Team Variance Attractiveness* as controls and *Photo* as baseline

Dep. Var.: Team Minimum Effort	<i>below-median</i>		<i>above-median</i>	
	(1)	(2)	(3)	(4)
Video Treatment	3.754** (1.720)	3.574* (1.794)	1.282 (1.885)	1.001 (1.830)
Homophily		14.049* (7.800)		-7.774 (9.511)
Team Cohesion Part 1		1.345 (1.188)		0.259 (1.194)
Team Risk Aversion		-0.506 (0.668)		-0.295 (0.749)
Team Attractiveness		-2.436 (4.081)		3.784 (3.614)
Constant	11.579*** (1.283)	8.066 (14.409)	13.333*** (1.295)	4.129 (20.840)
Observations	68	68	69	69

Robust standard errors clustered at the team level in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Linear regression of subgroups (below- and above-median in *Team Attractiveness*) with *Team Minimum Effort* in Round 1 as dependent variable, the *Team Attractiveness* as control and *Photo* as baseline

Dep. Var.: Team Minimum Effort	<i>below-median</i>		<i>above-median</i>	
	(1)	(2)	(3)	(4)
Video Treatment	1.458 (1.711)	1.984 (1.705)	4.012** (1.909)	3.960** (1.955)
Homophily		-0.991 (8.268)		2.819 (10.584)
Team Cohesion Part 1		1.584 (1.049)		-0.461 (1.216)
Team Risk Aversion		-0.596 (0.613)		0.198 (0.738)
Team Variance Attractiveness		-3.417 (11.027)		-0.207 (3.493)
Constant	14.167*** (1.154)	12.286 (8.975)	10.312*** (1.381)	9.205 (9.831)
Observations	68	68	69	69

Robust standard errors clustered at the team level in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Linear regression of subgroups (below- and above-median in *Team Variance Attractiveness*) with *Team Minimum Effort* in Round 1 as dependent variable, the *Team Variance Attractiveness* as control and *Photo* as baseline

C Communication Content Preparation and Analysis

Teams’ audios of the communication stage were recorded during the main study, and we informed subjects about this in the instructions. Immediately after the communication stage ended, a time window of a maximum of two minutes was available to upload the recordings to our server. Meanwhile, subjects could start reading the instructions for the next stage. Of the 204 teams that finished the study, we obtained complete audio recordings from 201 teams (*Audio* treatment: 65, *Photo*: 67, *Video*: 69). For three teams, the connection speed of all team members did not allow for the upload to complete within two minutes, so they had to be excluded from the communication analysis.

The audio recordings were transcribed using a local version of OpenAI’s Whisper (Radford et al., 2023), and speaker segmentation was done with Pyannote speaker-diarization (Bredin and Laurent, 2021). Afterwards, the transcripts were manually corrected.

For the analysis of communication content, we used several variables related to the total amount spoken in a team, the share of speech within a team, as well as the similarity of words used between teams. For the variable *Character Count*, which was used for the total communication length, we removed punctuation (as well as the speaker assignment) and counted the total number of characters used (excluding spaces). For the share of speech, we proceeded similarly by stripping the text for each speaker of its punctuation and counting the characters. The speech share of each team member was then calculated as the individual team members’ character count divided by the total character count of the team. At the team-level, we used the standard deviation of the speech share within a team (*Standard Deviation Speech Share*) for analysis.

In addition, we applied natural language processing to compare the communication content between teams and treatments. We followed data preparation and analysis recommendations by Ash and Hansen (2023), Gentzkow et al. (2019), and Werner and Andres (2024). Therefore, we first used standard approaches to clean texts by deleting double spaces, lowercasing, removing stopwords such as ‘the’ or ‘is’, and lemmatizing

to reduce words to their dictionary form (Ash and Hansen, 2023; Gentzkow et al., 2019). The resulting texts were, therefore, more comparable, while the relevant content was preserved. When analyzing data at team-level only, the speaker diarization was additionally deleted. As a first step, we looked at the 20 most frequent words for each treatment separately (see Figure 9). On first impression, the content seemed to be similar in all treatments, as in particular, the most frequent words were the same, and vacation-related topics were frequently discussed in all treatments. To go deeper into the content analysis, we conducted a cluster analysis and topic-modeling. Compared to topic-modeling, the cluster analysis works on a vector representation of our transcripts and provides a stricter separation of transcripts, as each is assigned to one specific cluster only (Ash and Hansen, 2023).

For the cluster analysis, we followed Werner and Andres (2024) and used agglomerative clustering with an euclidean metric, a type of Hierarchical Clustering for which clusters do not need to have a specific shape. The cleaned data was transformed into numerical form using the term-frequency-inverse-document-frequency (TF-IDF), handling each transcript as a document. Next, we excluded terms present in at least 99 % of documents (as they have little meaning when comparing transcripts), as well as very rare terms that are not present in at least 10 % of transcripts (sparsity threshold of 90 %). We compared the mean silhouette width (Rousseeuw, 1987) for different sparsity thresholds and number of clusters and used the elbow criteria to choose parameters. This led us to use the 90 % threshold in combination with three clusters (mean silhouette width = 0.11).²⁵ We found that all but three transcripts (two from the *Audio* and one from the *Video* treatment) were assigned to the same cluster, confirming our first impression that the transcripts were very similar in content.²⁶

For topic-modeling we chose two different approaches. The first approach uses Latent Dirichlet Allocation (LDA) (with Gensim (Řehůřek and Sojka, 2010)). First, a dictionary of tokens was built and further enriched by bigrams. Similarly to clustering, we removed tokens that were not used in at least 10 % of the transcripts or were in more than 99 % of the transcripts. This dictionary and the corresponding bag-of-words model (see Ash and Hansen (2023)) were used to train the LDA model. A comparison of C_v coherence scores (Röder et al., 2015) of different topic numbers was performed. The highest coherence score was reached when choosing two topics. The topics were then assigned to the individual transcripts, which showed that both topics were present in all of the transcripts. Looking at the most important words in each topic, we found that both topics were related to vacation. Therefore, based on this topic-modeling, we again could not detect any major differences between transcripts and, thereby, treatments.

As the transcripts were manually cleaned before analysis, we had an idea of the topics that were actually discussed. Most importantly, the discussion was resolved around vacations, and this icebreaker topic was mentioned in most of the teams. Still, we also identified other topics that we then used in our second topic-modeling approach.²⁷ We used the BART model (Lewis et al., 2019), specifically the pre-trained bart-large-mnli version, to perform text classification on the transcripts (not the cleaned data). For each speaker segment, we implemented the zero-shot classification (Yin et al., 2019) to check whether it matched any of the topics in our predefined topic list. We first looked at the overall *Topic Number* discussed in a team and then had a closer look at whether there are differences between treatments in the topic occurrence. The average *Topic Number* in a team was 8.17 in the *Audio*, 8.18 in the *Photo*, and 9.09 in the *Video* treatment. Pairwise comparisons with the *Video* treatment showed that the average *Topic Number* discussed in a team

²⁵A different number of clusters, lower sparsity thresholds, or different clustering methods did lower the mean silhouette width and were therefore discarded.

²⁶A similarly high difference in cluster sizes persisted also when increasing the number of clusters and changing the sparsity threshold.

²⁷Full list of topics used: vacation, politics, family, friends, career, work, sport, greeting, study, coordination, team, audio, economy, home.

was significantly higher in the *Video* treatment (Mann-Whitney U test: *Audio* vs. *Video*: $p = 0.004$, *Photo* vs. *Video*: $p = 0.002$). This is in line with the higher *Character Count* in the *Video* treatment and over all treatments the correlation between *Topic Number* and *Character Count* was positive and significant ($r = 0.21^{***}$). While most of the pre-defined topics occurred similarly often in all treatments (pairwise comparisons with Mann-Whitney U tests), there were some significant differences in the frequencies of occurrence of the topics “family” (*Video* > *Photo*: $p = 0.03$), “work” (*Photo* > *Audio*: $p = 0.05$; *Video* > *Audio*: $p = 0.003$), “study” (*Video* > *Audio*: $p = 0.085$; *Video* > *Photo*: $p = 0.006$), and “coordination” (*Video* > *Audio*: $p = 0.02$; *Video* > *Photo*: $p = 0.002$). These significant differences were in line with the overall higher number of topics discussed in the *Video* treatment. In addition, the *Photo* treatment was never significantly different from both the *Audio* and *Video* treatment. Overall, there was no clear pattern that could explain the observed differences in cooperation.

	(1) Audio	(2) Photo	(3) Video	(1) vs (2)	(1) vs (3)	(2) vs (3)
Character Count	3389 (653.44)	3528.15 (713.35)	3660.55 (633.68)	$p = 0.233$	$p = 0.007$	$p = 0.095$
Standard Deviation Speech Share	17.92 (9.43)	17.02 (9.26)	17.15 (8.89)	$p = 0.44$	$p = 0.702$	$p = 0.661$

Standard deviations are in parentheses. The last three columns report results from pairwise statistical comparison based on two-sided Mann-Whitney U tests.

Table 15: Summary statistics of team averages for Character Count and Standard Deviation Speech Share and pairwise treatment comparisons based on non-parametric tests

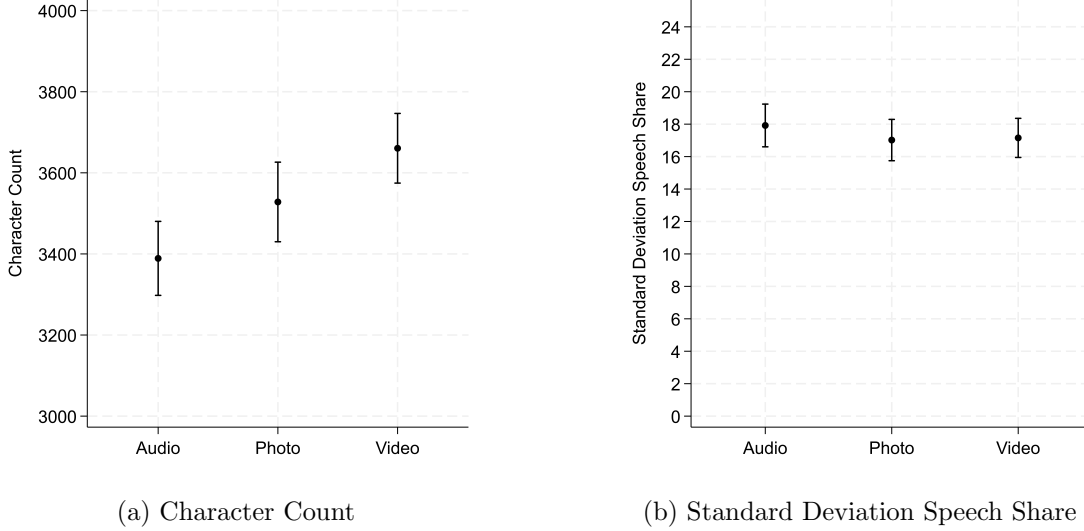


Figure 8: Character Count and Standard Deviation Speech Share by treatment

Dep. Var.: Team Minimum Round 1	(1)	(2)	(3)	(4)
Character Count	0.002** (0.001)		0.002** (0.001)	0.002** (0.001)
Standard Deviation Speech Share		0.061 (0.072)		0.057 (0.075)
Homophily			3.890 (5.102)	4.016 (5.144)
Team Cohesion Part 1			0.020 (0.783)	-0.010 (0.784)
Team Risk Aversion			-0.362 (0.437)	-0.338 (0.435)
Constant	6.670** (3.121)	13.067*** (1.270)	6.636 (6.360)	5.577 (6.362)
Observations	201	201	201	201

Robust standard errors clustered at the team level in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Linear regression with *Team Minimum Effort* in Round 1 as dependent variable, Character Count and Standard Deviation Speech Share as controls

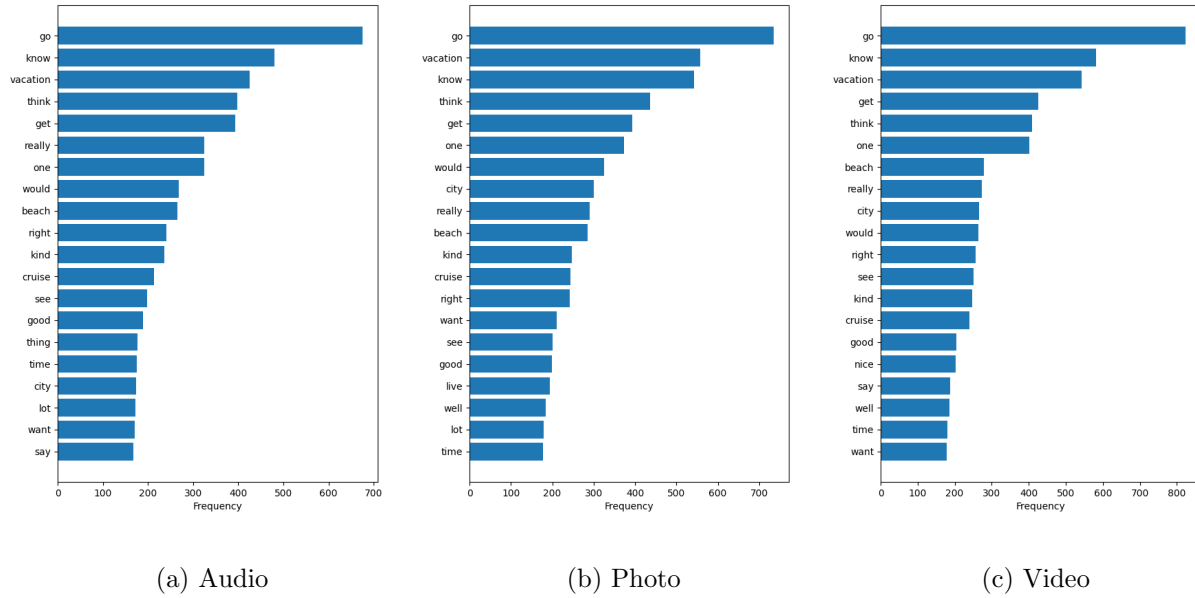


Figure 9: Top 20 words: Frequency by treatment