

Verbal interaction in a social dilemma

Zoë Adams^a, Agata Ludwiczak^b, Devyani Sharma^c, Magda Osman^d

^a Experimental Psychology, School of Biological and Chemical Sciences, Queen Mary, University of London, Mile End, London E1 4NS, United Kingdom
z.adams@qmul.ac.uk

^b Psychology and Counselling, School of Human Sciences, University of Greenwich, Old Royal Naval College, Park Row, London SE10 9LS, United Kingdom
a.ludwiczak@greenwich.ac.uk

^c Linguistics, School of Language Linguistics and Film, Queen Mary, University of London, Mile End, London E1 4NS, United Kingdom
d.sharma@qmul.ac.uk

^d Centre for Science and Policy, University of Cambridge, 10 Trumpington Street, Cambridge CB2 1QA, UK, United Kingdom
m.osman@jbs.cam.ac.uk

Corresponding author:

Dr. Zoë Adams
z.adams@qmul.ac.uk
Experimental Psychology,
School of Biological and Chemical Sciences,
Queen Mary, University of London,
Mile End, London E1 4NS.
07964 802 748

Declaration of interest: None

Funding source: This research was supported in part by the Economic and Social Research Council (ESRC) under project ES/R00787X/1: Rebuilding Macroeconomics, Putting in effort for the benefit of all: the role of reward and effort requirements.

Abstract

This article presents the first sociolinguistic examination of communication in a social dilemma. 90 participants (18 groups of 5) completed a modified public goods game with 2 rounds: an effort-based task and an unscripted online chat about the results. A linear regression shows that consensus-building language in the Round 1 chat affects cooperative behaviour in the Round 2 task. A qualitative analysis of 3 groups explores how participants use different recognisable styles of communication (registers) to strategically align with or disalign from one another (stancetaking). Each analysis is complemented with a quantitative visualisation of how (dis)alignment between participants unfolds in real-time. We found that successful groups employ registers associated with collective action, such as gameshow talk (“ouch. £69 out of a possible £120”) to encourage, punish, and pledge allegiance to one another. Less successful groups use registers that risk evoking mistrust and reducing obligation, such as business talk (“I approve”). We argue that a mixed methods approach to interaction and behaviour can reveal incremental shifts in consensus building that underpin quantitative outcomes.

Keywords: social dilemma, cooperation, verbal interaction, consensus, register, stancetaking

Introduction

Social dilemmas are a fundamental feature of human communication and cooperation. For example, the 2015 Paris Agreement seeks to slow global warming: ignoring the agreement would serve a country's immediate self-interests, but if every country fails to commit to their pledge, the planet suffers collectively. Although language and communication are central to these negotiations, there has been a notable absence of close linguistic analysis in social dilemma research. In this study, we demonstrate how a sociolinguistic approach can help illuminate key factors that underpin success or failure in the outcome of verbal coordination.

The analysis draws on data from a wider study exploring the effect of social forces and financial incentives on people's willingness to cooperate. The main experiment was a modified public goods game involving 90 participants who were distributed across 18 groups of five participants, six groups for each of three different conditions. Each group was given an opportunity to discuss their initial task performance in an unstructured format before repeating the task. We briefly report on the methods and main results of the full experiment before focusing our attention on the online chatroom data of three groups as case studies, one from each experimental condition.

Our analysis of these online verbal exchanges draws on two established concepts from the study of spoken interaction. The first is stancetaking—regarded as fundamental to communication (Jaffe 2009)—which concerns how speakers signal their evaluations of objects or ideas around them. People can use subtle verbal and non-verbal cues to communicate stances such as disapproval, authority, or solidarity. These cues have consequences for (dis)alignment with other individuals, ultimately facilitating or constraining cooperation. The second construct crucial to interaction is a higher level of sociolinguistic structure, namely register, i.e., how we speak in different social situations, for example, formally or in slang. Communication in all human societies is organised around frameworks of different recognisable styles known as registers. People can use these to adopt powerful 'voices' in conversation, regulated by rules of social convention and constrained agency. In the present data, we observe speakers employing a range of registers from parent-like roles to politicians' style to align with or disalign from each other. A close examination uncovers why these verbal acts might succeed under some conditions and not others.

We might ask why a classic behavioural economics experimental design should warrant a sociolinguistic analysis of verbal interaction in groups. Economics research is grounded in a quantitative and mathematical foundation that tends to steer clear of more qualitative questions that may seem too subjective (Lenger 2019). Sociolinguistics shares this positivist epistemology but, as its object of study is spoken interaction, it integrates quantitative generalisation with qualitative detail. When behavioural experiments involve communication, this qualitative dimension becomes very relevant. In our case, a mixed qualitative and quantitative approach to verbal interaction uncovers what is occurring incrementally within groups to achieve a group consensus outcome. We begin by outlining developments in research on the role of language in social dilemmas before introducing the notions of stancetaking and register, and how these sociolinguistic concepts can advance our understanding of collective action.

Theoretical background

Communication in social dilemmas

In social dilemmas, the rational strategy is to defect, so the high degree of cooperation observed in experiments has led to a substantial body of work aiming to understand people's motivations. Many economic studies attest to the importance of language; see Sally's (1995)

meta-analysis of 35 studies. Regarding medium, Bicchieri and Lev-On (2007) found that despite the absence of non-verbal cues, cooperation is still higher with computer-mediated communication (CMC) as compared to no communication. However, while face-to-face (FtF) communication results in greater levels of cooperation than CMC, Samuelson and Watrous-Rodriguez (2009) point out that it is nonetheless important to study the latter because technological advances are shifting many professional and personal interactions online, including real-life common resource allocation situations. In addition, they invoke Rice's (1984) observation that FtF communication makes it harder for people to deliberate or edit their responses, which may decrease the likelihood of objective reasoning and hinder cooperative efforts compared with CMC. Balliet (2010) highlights the further limitation of FtF communication being more costly and time-consuming. CMC therefore warrants our attention because it is often a preferred communicative choice or the only option available, particularly since the Covid-19 pandemic. Beyond medium of communication, there has also been extensive research on the timing of communication. Despite conflicting findings on the effect of communication before and during a social dilemma (Brosig et al. 2001; Frolich and Oppenheimer 1998), Balliet's (2010) meta-analysis of 45 studies revealed that there was no statistical difference between the two conditions.

A narrower body of research has explored how exactly group discussions affect cooperation. One account proposes that communication induces a norm of cooperation among participants who then believe that prosocial actions are expected (Bicchieri 2002). Another draws on Tajfel and Turner (1979), arguing that communication fosters a common bond, which creates an enhanced group identity (e.g., Orbell et al. 1988). In turn, members view themselves as part of the same in-group and are more inclined to cooperate. A final account posits that communication provides the opportunity for promise-making (Kerr and Kaufman-Gilliland 1994). To avoid cognitive dissonance, individuals aim to uphold their public commitments. In this case, cooperation results due to a personal norm rather than a social norm.

These approaches invoke largely social psychological accounts of the role of communication, which can leave unexplained why some instances of communication are more successful than others. For example, Torsvik et al. (2011) adopt a research design that examines different communication conditions (discussion, no discussion) but not the talk itself. Some economists have performed content analyses of discussions which occurred both prior to (e.g., Goren and Bornstein 2000; Kagel 2018) and during (McClung et al. 2017) decision-making in a social dilemma task. However, these studies overlook linguistic theory and stop short of analysing persuasion through unscripted talk. For instance, in identifying the main arguments for high levels of cooperation, Kagel (2018) includes the following example of the 'Most money' category from one participant: "Yo other team if you trust us we can both choose a (cooperate) and make some hashtag cash." The use of slang ("yo", "cash") and social media lexicon ("hashtag") deploys an informal register that could have been an attempt to reframe ingroup-outgroup boundaries and elicit cooperative behaviour. Its uptake would have depended on stances and registers asserted by the other participants, but these acts of conversational framing were not analysed despite potentially contributing the outcome.

Moreover, while previous studies make a valuable contribution by highlighting the potential of linguistic analysis for social dilemma research, statements have tended to be examined independent of their immediate communicative context. Kagel (2018) classified discussions into three categories: between opponent discussions (fairness, most money, threats), within team discussions (distrust, safety, retaliate) and messages following unilateral defection (no comment/mild upset, expectation of defection, expression of real upset/sarcasm). McClung et al. (2017) focus on shared intentionality, with discussions coded

into shared intentionality talk (reference to a shared goal, intentions required to achieve it, and mutual awareness of this common ground), individual goal talk (statements about a player's separate goals), task talk (statements about the task that did not reference shared or individual goals) or other talk. And Goren and Bornstein (2000) coded discussions according to six categories, including expressions of within-team mistrust and expressions of competitive intentions toward the out-group. The intersubjective dimension of cooperative exchange tends to be omitted in such approaches. For instance, McClung et al. (2017) found that shared intentionality statements increased costly helping and success, such as, "Have you checked (location in room)?" and "You help me, I'll help you." But we might ask, did the participant whose partner asked "Have you checked (location in room)?" respond, and if so, how? If there was no uptake, would this signal that intentions are not shared? How was the authoritative tone of "You help me" received by the partner? Lastly, McClung et al.'s (2017) category 'other talk' comprises filler statements, such as "Perfect", "Ok" and "That's fine", but the extent to which these agree with preceding statements varies, and may impact perceived willingness to cooperate.

We therefore aim to add established sociolinguistic tools for the analysis of talk to this body of work. The value of our interdisciplinary approach was first voiced half a century ago by the noted sociologist Erving Goffman following an exchange with prominent economist Thomas Schelling. Goffman (1961) remarked that the study of social games must consider interactional dynamics that often lack sufficient consideration in formal approaches to game theory. Several economists have since restated the case for a closer analysis of communication in social dilemmas. For example, Vollmer (2013) believes integrating microsociological investigations of everyday interaction with game theory offers analytical leverage in understanding coordination, while Sally (2002; 2005) refers specifically to the established notion of cheap talk:

"talk is not cheap, rather it is complicated, creative, implied, affective and effective, altering speaker and listener and any embedded game [...] once you let in a drip or two of literal meaning, you cannot prevent all of language from flooding in: poetry, promising, metaphor, irony, insult, intimacy, and the rest."
(Sally 2005: 263)

This kind of acknowledgement of the complex reality of spoken interaction is still rare in behavioural economics. The field has seen extensive adoption of game theory and rational choice theory in linguistics in recent years (e.g., Burnett 2019; Jäger 2008; Myers-Scotton and Bolonyai 2001), but the interface between purely rational calculations and social framing still remains under-examined. We show that this 'flood' of language is in fact a tightly structured signalling system that plays an important role in determining the extent to which cooperation is achievable in a given setting.

Stancetaking in communication

We adopt Du Bois's (2007) widely used model of stancetaking in this article. Du Bois defines stance as "a public act by a social actor, achieved dialogically through overt communicative means (language, gesture, and other symbolic forms)" (2007: 163). Speakers typically develop stances in response to those taken by other interlocutors rather than by a high-level prescribed norm. Norms for an interaction therefore develop dynamically and 'intersubjectively'—in response to preceding acts by others—within the context of wider, established rules for engagement. Du Bois (2007: 173) argues that stancetaking, the "smallest unit of social action", is always implicated in an utterance, as any choice of words reveals a

stance, e.g., the choice of *alleged* over *said*. His model combines two core concepts: evaluation and alignment.

Evaluation is the assignment of value to an object, either a referent external to discourse or an element of the discourse structure itself. For example, if a person says *what a great holiday!*, they are positively evaluating a discourse object external to the interaction, but if they say *obviously, I know that*, they are negatively evaluating the previous speaker's utterance or turn, an object within the conversational structure. In this way, we can evaluate anything, but from a social dilemma perspective, stance objects might include the reward (external to discourse) or a previous speaker's suggested strategy (discourse itself).

In evaluating an object, speakers position themselves along both epistemic and affective scales. For instance, *I know* is an epistemic evaluation that focuses on how certain an individual is about their assertion, whereas *I'm glad* is an affective evaluation, an expression of their emotional relation to the speaker or to a discourse object. Both scales are important in interaction. Jacknick and Avni (2017) argue that epistemic stancetaking is pivotal in anonymous scenarios particularly around high-stake topics. A social dilemma is such a scenario; indeed, our experiment entailed both financial and social risk. Defectors could win up to £70 if their entire group cooperated, but this would potentially expose them as selfish free riders. Cooperators could win far less if everyone else defected, but this would frame them as altruistic. Epistemic stancetaking in these situations allows an individual to establish credibility and assert expertise. Regarding emotion, neuroeconomic research attests to its important role in decision-making, for example irrational decision-making resulting from aversive affect (Engelmann and Fehr 2017).

The second component of Du Bois's model is alignment, which is the focus of our article. Any act of evaluation is typically also an act of alignment with or disalignment from other actors in the conversation. Figure 1 presents Du Bois's visualisation: As individuals evaluate objects, these evaluations position them relative to other evaluations in the discourse, which implicates inter-subjective alignment.

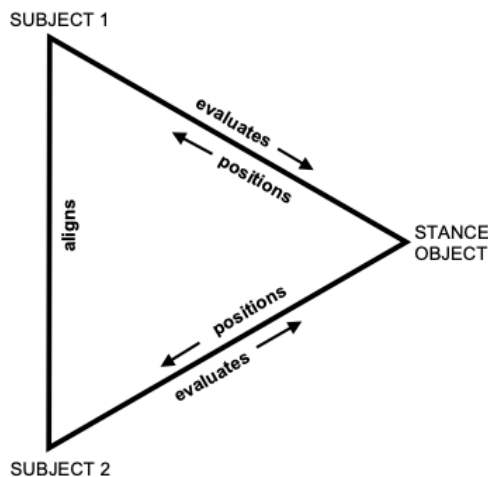


Figure 1. The Stance Triangle (adapted from Du Bois 2007, p. 163)

It is useful to distinguish between two aspects of alignment: evaluation and structural (see Kiesling et al. 2017 for further dimensions). Evaluation alignment indicates agreement with a speaker on their evaluation of a given object; this is paramount in a social dilemma, as cooperating is only better than defecting if everyone agrees on a strategy. An example is the

‘other talk’ category used by McClung et al. (2017), such as if a participant responds to a suggestion about a specific strategy from their partner with “Cool”, or “That’s fine”. Yet, evaluation alignment extends beyond shared intentionality, and other categories (e.g., in Kagel 2018 and Goren and Bornstein 2000): stance objects can include talk about the task, as well as talk about previous task results or talk unrelated to the task. This is not to underestimate the relevance of task-related communication (c.f. Bicchieri et al. 2011) but given that we communicate a stance with every utterance, from a sociolinguistic perspective, any aspect of a discussion in a social dilemma has the potential to help or hinder cooperative efforts. Structural alignment is an equally significant, but unexplored feature of previous content analyses in the social dilemma literature. This operates at a more conceptual level than evaluation alignment because it involves faithfully participating in the structure of an activity, which in our case is the online discussion. One example of structural alignment is providing an answer to a question. An instance of disalignment, however, would be if a participant asks their partner *are you willing to cooperate?* to which the partner responds *this experiment is so fun!* The failure to participate in the expected question-response structure introduces an element of disalignment. In the context of social dilemmas, structural alignment in such instances can be vital for maintaining group trust. A focus on alignment advances previous content analyses by moving away from isolated utterances towards how participants respond to one another, and the evolution of the interaction as a whole. Considering both structural and evaluation alignment allows us to observe how a speaker’s original stance comes to be endorsed, transformed, or undermined by the next speaker (Jaffe 2009).

It can be difficult to recognise whether or when a group has reached a decision because the process entails “incremental activities consisting of many minor steps” (Halvorsen 2015: 2). To accommodate intersubjective understanding being updated on a turn-by-turn basis (Kärkkäinen 2006), we track these increments systematically by following the methodology established by Kiesling et al. (2017). They attempt to overcome the intrinsic subjectivity of interactional research by embedding Du Bois’s categories within a computational implementation for measuring stancetaking, with corpus annotation and analysis. We follow Kiesling et al.’s use of a 1-5 scale to code online posts for stance elements, and additionally devise a format for visually tracking alignment via diverse stancetaking moves as they unfold among participants. Kiesling et al.’s nuanced treatment of alignment as continuous rather than dichotomous helps to show how participants maintain “strategic ambiguity” (Du Bois and Kärkkäinen 2012: 440)—a crucial device in a social dilemma game, given the trust and risks involved.

Frame and register in communication

As soon as an actor resolves to adopt a given stance during an interaction, they are faced with the problem of how to signal it. Stances are not rigidly tied to specific linguistic formulations. A person might express positive affect with *I’m delighted* or *yippee!* or any number of other alternatives. In each case, their choice of linguistic signal brings with it a conventionalised frame or ‘schema of interpretation’ (Goffman 1974), a socially shared set of principles for classifying and interpreting experience. The expression *yippee!* may invoke a playful frame, implicating childlike participant roles and reduced social distance. If a participant succeeds in imposing a given frame, it can limit other frames in the interaction and bind others to the appropriate behaviours for that frame.

Our linguistic choices thus involve the execution of recognisable social registers of speech (Agha 2005). A register is a particular type of language associated with specific social groups or speaker roles, for example news reporters, medical professionals, gangsters, and

innumerable other culturally conventionalised ways of being. Du Bois (2011) observes that analysing language does not just require a focus on the present moment of interaction *between* actors (stancetaking), but also exposure to discourses of the past *within* actors (register). The expression of stance thus relies on registers circulating in the wider social context (Jaffe 2009). Once an individual introduces these other shadow authors through their stylistic choices, their chosen stance is no longer theirs alone; the invoked ally—a parent, a confidante, a political leader—can consolidate the force of the stance (Coreen and Sandler 2014).

When we analyse stancetaking, then, we must additionally take into account the articulation of stances through recognisable voices, or personae or figures, in conversation. Even when actors believe they are simply being themselves, they are typically using a particular voice and framing that is suited to the situation. To date, economics research on social dilemmas has not examined how these universals of stancetaking and register lead actors to achieve or change their interactional goals.

Methodology

The social dilemma experiment

The data for this study come from a larger study exploring the impact of reputational information on decision-making in social dilemmas (Ludwiczak et al. submitted). The study is adapted from the “repeat-play” public goods game (e.g., Muehlbacher and Kirchler 2009) which comprises two rounds with a Task and Chat element as laid out in Figure 2. Informed consent was obtained for 90 participants (12 males, 78 females) who were recruited using opportunity sampling. The sample size was determined by previous social dilemma research which conducted group level comparisons (e.g., Goren and Bornstein 2000; Lev-On et al. 2010). Participants were all aged over 18 with no conditions that limited the use of their non-dominant hand. Participants were divided into groups of five, based on the order in which they were recruited. There were 18 groups, six groups for each of the three conditions.

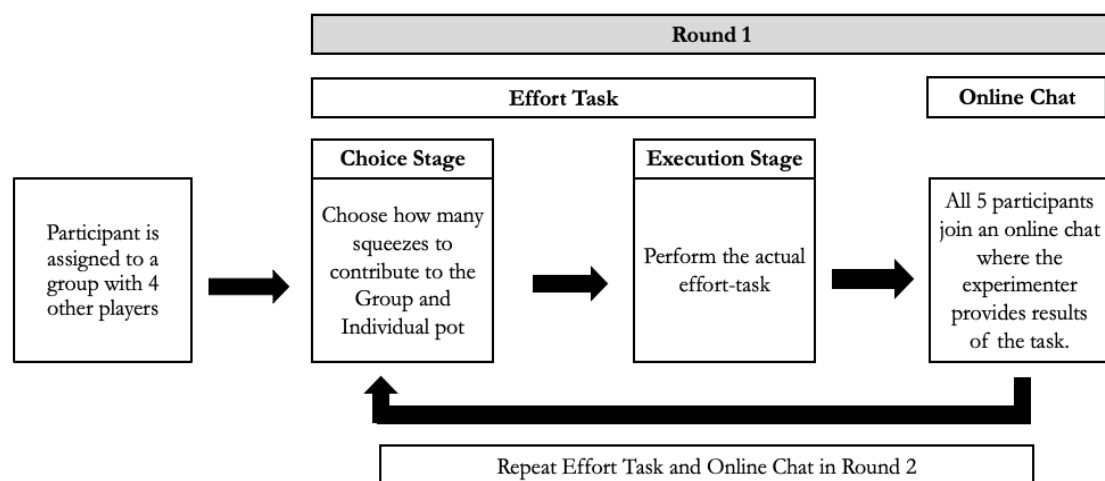


Figure 2. Schematic of experiment

After being assigned to a group with four others, participants completed the Effort Task which involved squeezing a hand-grip device 40 times. As this required effort, each squeeze was associated with a 40p reward. The reward resulting from each squeeze (i.e., effort) could either go to an individual pot or a group pot. Rewards from the individual pot

were given directly to the participant. The contributions that the participants made to the group pot were multiplied by 1.5 and divided evenly among all group members. A participant's final reward therefore depended on how many squeezes they contributed to their individual pot, as well as on how many squeezes all group members contributed to the group pot. Prior to squeezing the device (Execution Stage), they were asked to state how many squeezes they intended to contribute to the group pot (Choice Stage). For the Online Chat, approximately 2-6 days later, participants joined a platform called E-chat where the experimenter revealed the results. They had the opportunity to discuss their performance over a period of 36 hours. As the wider study tested hypotheses relating to feedback, reputation, and cooperation, this stage involved the crucial manipulation where the experimenter revealed three types of information depending on the condition: (1) intention – how many squeezes each participant intended to contribute to the group pot, and the group pot total; (2) action – how many squeezes each participant actually contributed to the group pot, and the group pot total; (3) baseline – the group pot total. Participants were required to post at least one task-related comment. In Round 2, they repeated the Effort Task on average 4.8 days later, and the experimenter ensured that they had read the entire online discussion before doing so. They then took part in another Online Chat. Participants could earn up to £70 (\$92), but, for ethical purposes, all participants received a minimum of £40 (\$53) to avoid vast discrepancies in payment.

The present study briefly reports on the overall findings, but then focuses on the Online Chat of Round 1, and in places Round 2, of 3 groups. First, we turn our attention to the construction of the alignment visualisation.

Coding and annotation

For the coding of alignment, we closely follow Kiesling et al.'s (2017) procedure, which begins with identifying the object of evaluation, or “stance focus”, for each utterance in each discussion. If a comment included multiple utterances by a participant but no single stance focus for the entire comment, then the comment was split according to stance focus, as in (1). (Subject IDs take the form of group name, subject number, gender.) Equally, if there were multiple consecutive comments with the same stance focus, then the comments were joined as in (2). This ensured that each annotation was linked to one stance focus. Greetings were excluded or, if followed by another comment, they were merged as in (3).

- (1) **Green5F** (22:49): This looks a really interesting study! (*Focus: study*) I put in all of my 40 in the group pot in the first round and will do the same in the second round 😊 (*Focus: contributing all 40 to the group*)
- (2) **Gold5M** (10:31) Hi everyone, that's a good amount! The more we contribute the more we will get since it's x1.5 (*Focus: Round 2 strategy*)
Gold5M (10:36) Shall we go all in for part 3? If we all do, instead of £16 we each get £24 (*Focus: Round 2 strategy*)
- (3) **Orange4F** (08:31): Hi everyone, sorry for coming late to the party! It's interesting seeing everyone's contributions, particularly Orange3F – I didn't expect anyone to be that generous (*Focus: Orange3F's contribution*)

Each utterance was then annotated for a value of 1-5 for alignment (the same number of distinctions used in Kiesling et al. 2017). A score of 5 indicates high alignment, 1 indicates low alignment. The first and third author double-coded the first discussion blind and

compared their results (Cohen’s kappa for inter-rater reliability: 0.654); disagreements were examined to develop more explicit heuristics for ambiguous cases. The first author then proceeded with coding discussions two and three. Table 1 displays examples of utterances that received an alignment score of 1 and 5. The utterances are listed along with the preceding comment, which helps to clarify the alignment scores. Although we do not analyse evaluative polarity here, the stance focus in the final column shows how alignment emerges in part through evaluation of stance foci.

Table 1. Examples of alignment coding

Ex.	Username	Comment	Alignment	Stance focus
1	Gold4M	we can do around 25-30 in the pot		Strategy for Round 2
	Gold2M	why not maximize it and put all 40 in the pot? bigger the contribution bigger the reward don't you think?	1	Gold4M’s suggestion on strategy for Round 2
2	Green4F	I’m happy to add all credits to the group if everyone is in agreement		Strategy for Round 2
	Green1M	Yay!!	5	Green4F’s promise to cooperate

Kiesling et al. (2017) did not annotate the first posts for alignment in their data, as they were initial in the interaction. By contrast, in the present data the first post was always the experimenter, who revealed the results from the task in Round 1. As the experimenter’s comment always implicated the entire group’s behaviour, the following comment—the first by an experimental subject—was always aligning with or disaligning from the group to some degree. For example, in (4), we can see that Gold1F, who commented first, is aligning with her group by positively evaluating the performance of all members, albeit to a slightly weak extent as suggested by the qualification of “too”.

(4) **Gold1F** (09:00): Everyone’s contributions are not too low! Which is a good thing.

Based on previous studies (e.g., Goren and Bornstein 2000; Kagel 2018; McClung et al. 2017), we start with a summary which shows that consensus-building discourse in our data does correlate overall with an increase in cooperative behaviour. This complicates proposals which assert that communication per se increases cooperation, for social or psychological reasons. Under that view, we should not expect to see a correspondence between the quality of the interaction and the level of success in a collective action outcome.

Findings

To see whether degree of consensus within communication impacts later behaviour, we explored the effect of online discussions on levels of cooperation for all 18 groups. The key independent variable here is the consensus score for each group in the Online Chat of Round 1. We used an aggregate score for consensus talk that was distinct from the alignment analysis presented in the next section. Consensus talk examines statements in isolation while, as mentioned previously, alignment is based on the dynamic relationship between utterances. Moreover, our analysis of consensus talk focused specifically on how each statement supported a group consensus to cooperate, that is contribute all 40 squeezes to the group pot. We therefore identified all consensus-building comments in the transcripts and assigned each one a value between -1 to 1, in increments of 0.25. For example, one participant agreed with

a previous participant about the ease of the task: “Yeah it was pretty quick!” This statement received a 0 score for consensus, because it neither facilitated nor hindered a group decision to contribute all 40 squeezes to the group pot. However, as this participant was expressing agreement with the previous speaker regarding their evaluation of the stance object (the task), the statement would receive a score of 4 for alignment (rather than 5, as the evaluation was somewhat mitigated by the adverbial “pretty”).

This numerical assignment of -1 to 1 also offers a more granular analysis of the relationship between talk and cooperation compared with previous studies (e.g., Goren and Bornstein 2000; Kagel 2018; McClung et al. 2017), which assigned each statement to a category. The first step of our coding process involved coding each utterance according to four categories: promises, questions, imperatives, and comments. This allowed us to develop minor categories and delve deeper than the categories in other studies, such as Goren and Bornstein’s (2000) category (c) “Explicit willingness to signal cooperative intentions to the out-group by lowering in-group contribution”. For example, our promises category was broken down into: (1) explicit specific (e.g., “Yup, *I’m* happy to put in *all 40*”) which received a score of 1; (2) conditional specific (e.g., “*If* we all give *40* squeezes to the group everyone will earn more”) which earned a score of 0.75; (3) explicit vague (e.g., “Agreed. *I’ll* give *more* next time”) which received a score of 0.5; (4) conditional vague (e.g., “*If* everyone puts into the group pot earnings will be *more*”) which received a score of 0.5; or (5) explicit conditional specific (e.g., “*I’m* happy to contribute *all 40* if everyone else is”) which received a score of 0.5. It should be noted that as these were promises, they all received a positive score. This coding scheme also builds on McClung et al.’s (2017) shared intentionality but probes further. For example, one minor category was called ‘Planning shared goals prior to hunt’ and contained utterances, such as “We should/could collaborate”, “We should/could work as a team”, “We should/could trade screws”. Conflating “should” and “could” minimises an important difference because the former implies obligation, while the latter only signals possibility. Otherwise put, while we capture concepts such as shared intentionality and willingness to signal cooperative intentions, our coding scheme also studied how the extent of cooperative behaviour is sensitive to linguistic nuances.

Inter-rater reliability was tested via blind double coding of 5% of the data: Agreement within 1 level out of 9 was at 75% before discussion and 85% after. Table 2 shows the difference in group squeezes between Round 1 and 2 and consensus score for each group which is visualised in Figure 3. The difference in squeezes ranged from -16 to 115 ($M = 47$, $SD = 34.09$), and the consensus score ranged from -0.039 to 0.696 ($M = 0.47$, $SD = 0.24$). Having assigned each group an average consensus score, we conducted a linear regression analysis using the *lm* function in R (Bates et al. 2015). To test for an effect of communication on levels of cooperative behaviour, the difference in group squeezes was taken as the dependent variable, whereby a higher score indicated a greater degree of cooperation. Results revealed a significant effect of consensus score on cooperation ($\beta = 92.54$, $SE = 26.74$, $F(1,16) = 11.98$, $p = .003$) and a correlation coefficient of 0.67.

Table 2. Consensus score and the difference in group squeezes between Round 1 and Round 2 for each group

Group	Differences in squeezes	Consensus score
Mauve	115	0.696
Green	85	0.769
Yellow	82	0.675
Rose	77	0.367
Blue	70	0.800
Purple	65	0.350
Navy	57	0.667
Orange	53	0.528
Violet	46	0.600
White	44	0.357
Grey	43	0.500
Gold	42	0.734
Plum	34	0.306
Red	32	0.292
Silver	30	0.050
Black	2	0.500
Turquoise	-15	0.286
Scarlet	-16	-0.039

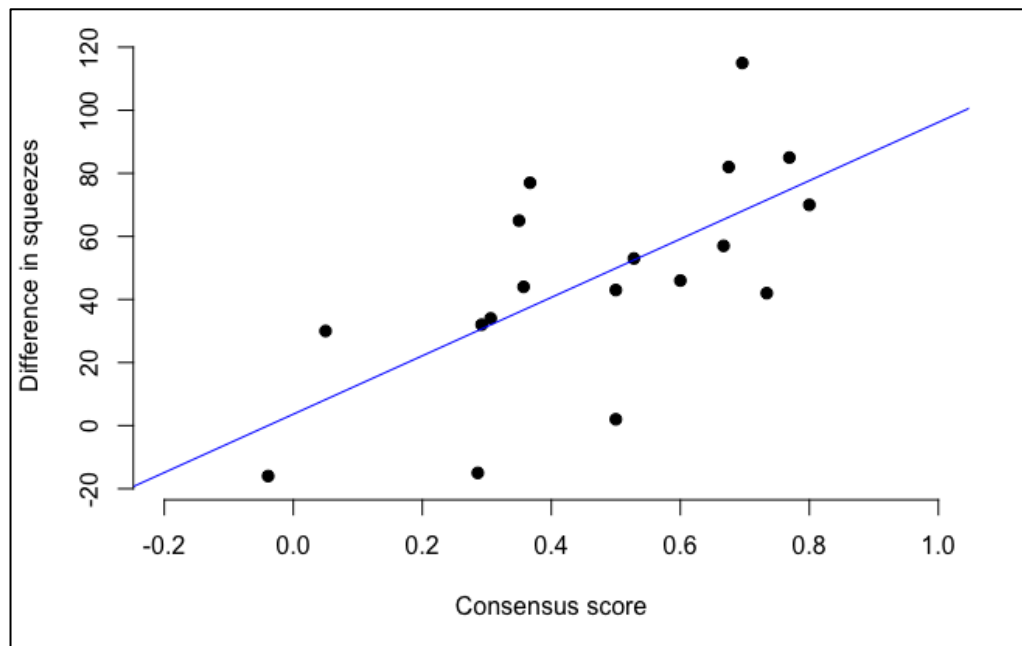


Figure 3. Consensus score and the difference in squeezes between Round 1 and Round 2 for all 18 groups

The results corroborate previous suggestions that it is not simply communication that influences cooperative behaviour, but the quality of communication (e.g., Kagel 2018; McClung et al. 2017; Torsvik et al. 2011). However, this leaves a further question unanswered: *why* did some discussions move towards more cooperative talk than others? The consensus analysis cannot capture the intersubjectivity of the discussions (stancetaking) or the contribution of different social voices (registers), known to permeate all interactions. In the case studies that follow, we therefore conduct a close analysis of the strategies participants use to achieve cooperation through talk, and why they meet with variable success. One group from each condition, marked in bold in Table 2, was selected for a mixed methods analysis combining qualitative and quantitative tracking of discourse: Orange (action), Gold (intention), and Green (baseline). The discussions, which varied in length from 13-15 comments, were selected because they capture a range of degrees of success in achieving collective behaviour and provide a rich array of interactional styles to examine. We close by briefly drawing parallels between the linguistic behaviour observed in these three groups and those in other groups, particularly Turquoise, Scarlet and Mauve.

For each group, we first present a summary of how many squeezes each participant contributed to the group pot out of 40 in the Effort Task of Rounds 1 and 2 (Tables 3-5), and then analyse the Online Chat in Round 1. This is presented alongside an alignment graph tracking the strength of alignment of each comment in relation to the previous participant's comment (Figures 5, 7, 9). This delves into the incremental shifts in interpersonal alignment that partly underpin the composite consensus scores in Table 2. The transcripts exclude the first post, in which the experimenter reveals the results, as the format of those posts always comprised a greeting to the group, and then the following: Condition 1 (intention): "Username[1-n] chose to contribute X squeezes to the group pot" and the group win; Condition 2 (action): "Username[1-n] contributed X squeezes to the group pot" and the group win; Condition 3 (baseline) "The group win for Round 1 is £X".

Green team: The family gameshow

Table 3 shows that the Green team (baseline) was extremely successful in increasing their cooperative behaviour following their discussion. After the Online Chat, every participant increased their contributions to the group pot to the maximum number of squeezes.

Figure 4 provides the transcript of this group's discussion and Figure 5 presents the alignment visualisation. Our analysis centres around the interactional frame created by the first commenter, Green1M, and the resulting tension and resolution in the chat.

Table 3. Squeezes contributed to the group pot by the Green team in Round 1 and Round 2

User	Round 1		Round 2
Green1M	40	Discussion	40
Green2M	0		40
Green3F	15		40
Green4F	20		40
Green5F	40		40
Total	115		200

1	green1M (13:52) Hi Peeps! Hope you all acted altruistically and weren't greedy!! Scores on the doors tomorrow!! :-O Results revealed: Group win
2	green2M (09:08) Hey guys! green2M (09:09) thats a decent group score, lets use it as motivation to get a bigger pot
3-4	green1M (09:41) Ouch. £69 out of a possible £120. // Come on people, add all the credits to the group pot and we get 1.5x the amount for individuals. If we all do the same, we will end up better off
5	green3F (10:19) Hi all, that's an interesting figure
6	green3F (10:19) The task was surprisingly quick and easy. Not sure what I was expecting though
7	green2M (11:17) I agree with c4r9 we should all think of the group in order to get a bigger amount.
8	green4F (14:11) I'm happy to add all credits to the group pot if everyone is in agreement.
9	green1M (15:53) Yay!! green1M (15:53) All for one and one for all!! ...£££!!!
10	green5F (22:46) Hi everyone, hope you're all well green5F (22:48) Really glad to hear everyone getting on board with putting all their squeezes in to the group pot
11-12	green5F (22:49) This looks a really interesting study! // I put in all of my 40 in the group pot in the first round and will do the same in the second round 😊
13	green5F (22:50) I hope we can make it to the £120 in the second round! Come on everyone ... let's band together and get the best possible outcome

Figure 4. Transcript of discussion for group Green. Sequential comments with the same stance focus by a single participant are assigned one alignment score. Comments separated by ‘//’ have a different stance focus and are coded separately.

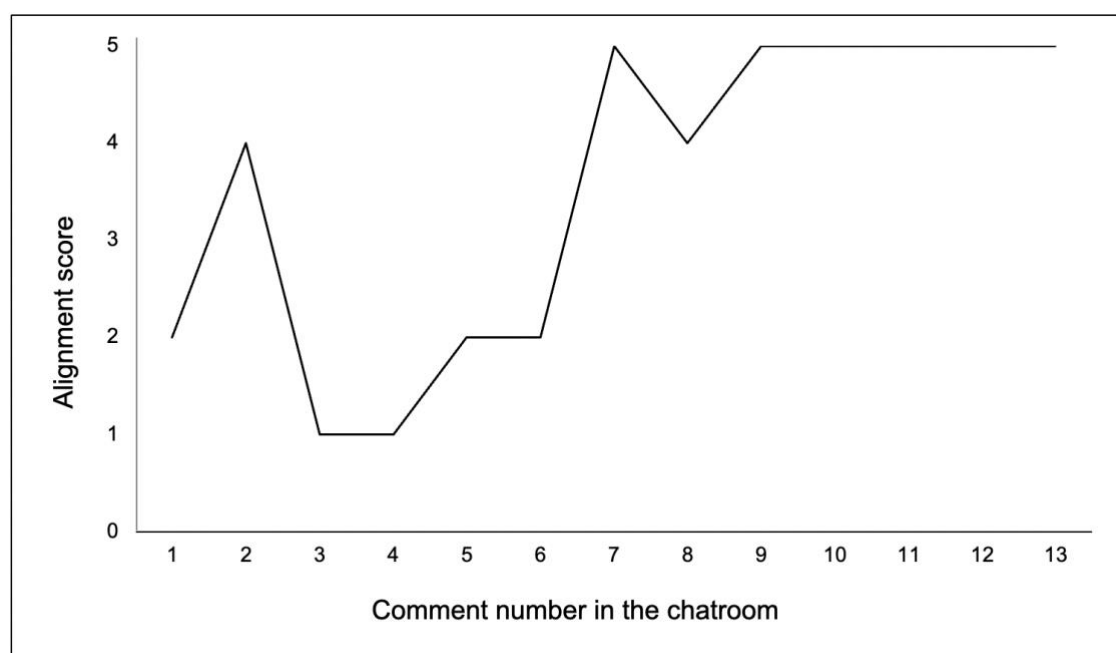


Figure 5. Alignment score for each comment in relation to prior speaker in Green team discussion (1 = disaligned, 5 = aligned)

In his opening comment, Green1M first reduces social distance with a playful solidarity marker (“Peeps!”), then delivers a reputational threat. The use of morally evaluative terms such as “altruistic” and “greedy” without further justification and little verbal hedging construct an authoritative role (van Leeuwen 2007), even before the result is announced. Green1M maintains this paternalistic role with the framing device “scores on the doors”, popularised by Bruce Forsyth who hosted an inter-generational family contest called The Generation Game. The use of ‘scores’, a feature of any game, combined with this rhyming pattern creates a light-hearted tone. However, he embroiders this framing device with ambiguously playful signals: double exclamation marks intensify the emotive force of the moral sanction (Vandergriff 2013) but can also seem comically hyperbolic; the emoticon hints at his own vulnerability for trusting his group but also issues a warning to others of the big reveal of their moral compass. In this way, he comes across as the family judge presiding over everyone’s actions (Ochs and Taylor 1995).

This compact but powerful package of signals invokes a domestic disciplinary frame and potentially constrains responses available to others. Recall the link between structural (dis)alignment and evaluation of discourse objects: if participants risk disaligning from Green1M structurally by failing to acknowledge his comment, Green1M’s evaluation—in favour of the collective strategy—might also have been less valued. As it turns out, Green1M is wholly successful in bringing the group around to his evaluation. How does he achieve this?

The first factor is that his main ‘rival’, Green2M, adopts a conciliatory strategy. Green2M is the only player to contribute 0 and so is a direct target of Green1M’s threat (though only he is aware of this). In comment 2, he avoids the risk of losing face by opting for a positive tone and placating Green1M: he praises the group result, echoes Green1M’s gameshow term “score” (not used in any of the other 18 groups), and uses the imperative “let’s” to share in Green1M’s authority.

Instead of aligning with Green2M, Green1M escalates the face threat to Green2M and the group by conveying the full force of personal injury with “Ouch.” As the full stop is optional in texting, its presence can express subtle interpersonal information (Houghton et al. 2018), in this case dismay through a move away from his high-affect exclamation marks. The gameshow-esque “£69 out of a possible £120”, which is again in score format, highlights the magnitude of his disappointment. These explicit signals of disappointment are an act of social shunning, which function to offset the perceived short-term benefits of uncooperative behaviour and reduce the incentive of freeriding (Gachter and Fehr 1999). The parental tone (imperatives, exhorting good behaviour, lack of mitigating politeness) reintroduces the original status imbalance and positions the group’s problematic behaviour as the focus of his negative evaluation. Green2M again repairs potential damage in comment 7 by seizing the opportunity to rehabilitate himself positionally and aligning explicitly with Green1M.

A second overall factor facilitating Green1M’s success is his artful double-voicing of authority (parental/gameshow frame) and solidarity (socialist frame), softening the unmitigated format of the former stances with humour and positive affect in the latter. As support for his stance starts to be voiced, Green1M quickly abandons his risky stern register and adopts almost the reverse register, a childlike “Yay!!” followed by a collectivist call to arms: “All for one and one for all!!...£££!!!”, with a return of an excitable tenor through punctuation. Even his vacillations in mood are in line with the emotive nature of gameshows which “excite passions” as contestants “weep and dance, kiss and hug, scream and shout” (Roe et al. 1996: 49). More than a jumble of stances, the combination of Green1M’s different moods can be viewed as a single performance. This allows us to see that, despite the risks involved, such displays of emotion can spark social action (Du Bois and Kärkkäinen 2012).

The final factor is Green5F's closing confirmation. Her positive evaluation with an asserted stance object of "everyone getting on board" implicates the entire group, including Green3F who structurally disaligned from Green1M in comments 5-6. She places on record her full cooperation in Round 1, exempting herself from Green1M's earlier punishment, and structurally mirrors Green1M's animated gameshow affect with a smiley emoji, exclamation marks, his phrase "come on", the gameshow term "round" (only featured two other times in the entire dataset), and the communal phrase "band together" (Twenge et al. 2012).

Comments made in the Online Chat of Round 2 are also revealing. Drawing on Goffman (1956) and Brown and Levinson (1987), psychologists Dunning et al. (2014: 124) argue that participants in social dilemmas cooperate because it is insulting to withhold trust. In other words, they honour an injunctive norm to trust "not because it is what they want to do but because they feel it is an obligation of their current social role." Establishing trust was hardest in the Green team because they were in the baseline condition where past individual contributions were unavailable. When the experimenter revealed that everyone cooperated in Round 2, Green4F said: "Wow great job everyone, no one betrayed the group 😊!" She trusted the group on the behavioural level, as indicated by her group contributions, but her surprise suggests that she did not trust her group at the cognitive level, along the lines of Dunning et al.'s proposal. However, while this doubt shows that cheap talk is not reassuring (Kydd 2000), even in the absence of a costly signal, such as past behaviour, Green4F seemed obliged to cooperate which aligns with Green1M's parental and socialist registers.

Taken together, despite a rocky beginning, the discussion ends with a high degree of alignment and full cooperation in Round 2. We see how an agent can make a chosen norm salient to induce cooperative behaviour (Bicchieri 2002; Bicchieri and Lev-On 2007), leading to structural and evaluative uptake in the discussion. Green1M created a group identity (Orbell et al. 1988) through registers associated with cooperation and the discussion shows how trust can be fostered at zero acquaintance for the greater good (Dunning et al. 2014). As Du Bois (2007: 173) observes: "we care about the state of the game, too: how it is played, who plays it well and fairly, in what condition the players leave the turf."

Orange team: The politician

Table 4 shows that the Orange team (action) was not very successful in increasing their cooperative behaviour following their discussion. Although the number of squeezes contributed to the group pot increased, only two participants contributed the maximum amount (Orange1F, Orange3F), only one changed their behaviour in this direction, and two failed to increase their contributions at all (Orange4F, Orange5F). As we will see, this group is less successful in aligning during their interaction as well.

Figure 6 provides the transcript of this group's discussion, and Figure 7 provides the alignment visualisation. In Figure 7, we see a steadily declining level of alignment, which does not bode well for an outcome of increased collective behaviour. The core difficulty appears to be an unresolved tension in register choice (potentially arising out of underlying stance differences) and lack of harmonisation between Orange3F and Orange4F, which impedes trust and coordination overall.

Table 4. Squeezes contributed to the group pot by the Orange team in Round 1 and Round 2

User	Round 1		Round 3
Orange1F	15	Discussion	40
Orange2M	0		29
Orange3F	40		40
Orange4F	20		19
Orange5F	4		4
Total	79		132

1	orange1F (09:13) I think if we give half to the group it may make the reward higher but only by a little bit
2	orange1F (11:19) actually I think it would be better if we all gave at least 30 to the group pot since we would all benefit from it
3	orange2M (12:12) 30 seems like a good number
4	orange3F (21:33) Hello everyone, as you can see I have given all of my squeezes to the group. I believe we should all do the same in the next stage in order to maximise our income
	orange3F (21:33) Is anyone else of the same opinion as me? Logically none of you have to lose but only gain.
5-6	orange2M (22:27) Thank you for the generosity orange3F // For sure, if we all do it, it will benefit all of us to a great extent
7	orange4F (08:31) Hi everyone, sorry for coming late to the party! It's interesting seeing everyone's contributions, particularly orange3F - I didn't expect anyone to be that generous
8	orange4F (08:33) We all gain if we all put the total amount in, but given this is anonymous and we don't know what everyone else to do, it's hard to decide
9	orange3F (11:59) I believe we could trust in the human decency of each other and leave the experiment with both money and a satisfied feeling of cooperation !
10	orange3F (12:04) I will be doing the same as in the previous part and hope u can retaliate my blind trust!
11	orange1F (12:09) I'm more than happy to give all my squeezes to the group as we'll all win from it 😊
12	orange4F (18:23) Get rich or die trying!
13	orange5F (22:38) I think that sounds like a good plan - if we all contribute at least something to the group pot.

Figure 6. Transcript of discussion for team Orange. Sequential comments with the same stance focus by a single participant are assigned one alignment score. Comments separated by ‘//’ have a different stance focus and are coded separately.

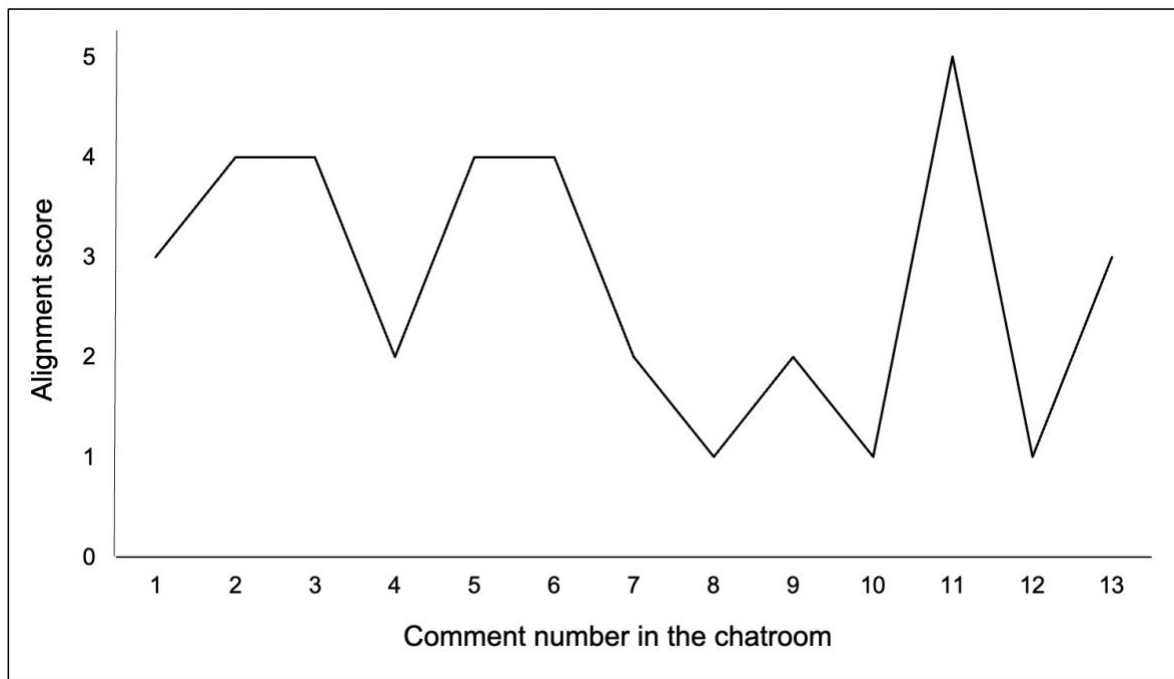


Figure 7. Alignment score for each comment in relation to prior speaker in team Orange discussion (1 = disaligned, 5 = aligned)

The discussion begins with Orange1F and Orange2M agreeing to increase their contributions. The alignment slope in Figure 7 starts to decline when Orange3F tries to encourage full cooperation and Orange4F expresses distrust. This triggers a power play between the two. Unfortunately, despite a highly agreeable comment by Orange1F in comment 11, the discussion does not recover and Orange5F's final non-committal promise leaves the discussion on an uncertain note.

Let us begin with the political tone of Orange3F. Although Orange1F and Orange2M did not greet the group, Orange3F joins the discussion with a formal greeting in comment 4 immediately followed by a disaligned stance in favour of full cooperation. She draws attention to her cooperative behaviour in Round 1, emphasising others' lack of altruism through her use of "you". A number of features of her discourse channel political speech. The present perfect "have given" operates as a persuasive technique in the realm of politics, linking past decisions to the present moment (Fetzer and Bull 2012), as does the strong epistemic certainty of "I believe" (Fetzer 2014). This contrasts with the lower epistemic commitment in earlier comments, in forms such as "I think", "may", and other hedging devices (Milkovich and Sitarica 2017). Orange3F uses the inclusive pronoun in "we should" to momentarily align with her group and present her goals as the audience's goals, another tool used by the "inspiring orator" of politics (Joseph 2006: 13). However, when subsequently eliciting her group's thoughts, she does not opt for a low social-distance format such as 'What do you think?', but rather asks if anyone shares her opinion, positioning it as the norm. The clause that follows ("Logically...") again strongly asserts epistemic authority and discourages disagreement. The formal, authoritative, and somewhat argumentative nature of her comment is typical of political discourse (Archakis and Tsakona 2010; Van der Valk 2003; Vukovic 2014). Framing her comment as a speech, she plunges the group into the role of an audience (Jaffe 2009). As with Green1M in the first case study, Orange3F places herself in a risky position with this bold framing choice. She positions herself as an authority, but lacks the nuance of Green1M's frame-shifts.

Although Orange3F's register ultimately fails to unite, it influences the format of later comments. Orange2M aligns with Orange3F in a tone of equal formality. The content of his comment, however, is not entirely aligned: the inclusion of "if we all do it" reiterates the risk. This is reinforced by Orange4F who first disaligns from Orange2M by labelling the results as "interesting" and then offers a slightly critical assessment of Orange3F's generosity. Orange4F maintains an uncertain stance to Orange3F's proposal, explicitly introducing concerns over trust. In this unsettled, low alignment mood, Orange3F attempts to regain control (comments 9 and 10) with a moral argument to invoke cooperation by making an emotional appeal to their "human decency", "hope", and "cooperation", still in the political oratorical register (Duranti 2006) but with a note of frustration in the punctuation. Her use of netspeak "u" and awkward, possibly non-native use of "retaliate" is somewhat incongruent with her chosen register, risking her credibility. Goffman (1956: 33) remarks our tendency to "pounce on trifling flaws as a sign that the whole show is false" and this "forces an acutely embarrassing wedge between the official projection and reality."

In a final blow to Orange3F's authority, Orange4F upends Orange3F's imposed frame with a phrase popularised by a hip-hop album in the 2000s entitled *Get Rich or Die Tryin'*. Like Green1M's direct quotation from a gameshow, the reference here is unambiguous, in particular its stark contrast to the political register in play: hip-hop symbolises resistance against status-quo politics (Perry 2004). Her surface sentiment is positive, but the structural disalignment, including lack of direct relevance, from Orange3F is painfully evident. Orange3F is attempting to resist her mistrust, but Orange4F responds with mockery and defiance. By contrast, Orange1F does perform a "stance follow" (Du Bois 2007: 161), i.e., takes up the action made relevant by Orange3F, agreeing to contribute all 40 squeezes to the group and executing this promise in Round 2. This is the only moment in the discussion where alignment, evaluation and affect are all high. The conversation ends with Orange5F agreeing to cooperate, swiftly followed by a reinterpretation of Orange3F's strategy as "contributing at least something". Indeed, Orange5F contributed only 4 squeezes in Round 2, as in Round 1.

In sum, it is important to reiterate that the Orange Team was in the action condition, and so the authenticity of each pledge was available for scrutiny alongside individual contributions in Round 1 (see also Gold team below). Orange2M and Orange1F sought enough reassurance to cooperate, offering pledges following Orange3F's speech. However, their past behaviour is visibly inconsistent with their pledges to cooperate, raising legitimate concerns about defection. Not only this but trust issues, notably expressed by Orange4F, were triggered when Orange3F awkwardly attempted to impose her stance through a political voice. In fact, studies have found that discussions of mistrust are absent in cooperative groups (Goren and Bornstein 2000; Majeski and Fricks 1995). This is in line with Majeski and Fricks (1995: 629) who warn that "saying something in the wrong way (...) can be worse than not talking at all and might harden group-based distrust." Such an interpretation receives support in Round 2 when, upon discovering the cooperative behaviour of her group, Orange4F remarks: "wow, I guess I need to be more trusting". Therefore, while the Green team are evidence that cooperation is not contingent on costly signaling, here we see that costly signaling, by Orange3F who contributed 40 squeezes in Round 1, can be negated by consequent low-quality interactions. Indeed, Bicchieri and Lev-On (2007) argue that in the impoverished online chatroom environment, without auditory or visual cues, trust is harder to establish because participants are more removed from settings where promises are made and kept.

Gold team: The business meeting

The Gold team (intention) achieved only limited overall success in increasing their cooperative behaviour following their discussion. Table 5 shows three participants did shift fully to cooperative behaviour in Round 2, but two defected (Gold1F, Gold5M). In this final case, we highlight some limitations of chat data, particularly due to sincerity violations.

Figure 8 provides the transcript of this group's discussion, and Figure 9 presents the alignment visualisation. In this case, the discussion is characterised by equivocation and ultimately defection by two players, some dissonance in alignment and negotiation, and subtly contrasting discourse styles.

Table 5. Squeezes contributed to the group pot by the Gold team in Round 1 and Round 2

User	Round 1	Discussion	Round 2
Gold1F	12		5
Gold2M	15		40
Gold3F	20		40
Gold4M	6		40
Gold5M	30		0
Total	83		125

1	gold1F (09:00) Everyone's contributions are not too low! Which is a good thing.
2	gold2M (09:11) That's a decent amount of contribution!
3	gold2M (09:12) The more everyone contributes the more you can all get
4	gold3F (09:24) Well done group 😊
5	gold1F (09:28) The more we contribute, it benefits us all
6	gold2M (09:32) Deffo will ensure group pot increases in Part 3
7	gold4M (10:06) Yes we can increase the contribution as it is quite profitable to us all.
8	gold5M (10:31) Hi everyone, that's a good amount! The more we contribute the more we will get since it's x1.5
	gold5M (10:36) Shall we go all in for part 3? If we all do, instead of £16 we each get £24
9	gold2M (10:40) Best possible solution, we all go in for the pot and it becomes larger
10	gold5M (10:42) Exactly
11	gold4M (13:35) we can do around 25-30 in the pot
12	gold2M (13:39) why not maximize it and put all 40 in the pot? bigger the contribution bigger the reward don't you think?
13	gold5M (16:25) Yea I agree. Put all 40 in the group pot and we'll have a bigger reward
14	gold2M (16:39) I approve ^
15	gold4M (13:25) ok then i will put all 40

Figure 8. Transcript of discussion for group Gold. Sequential comments with the same stance focus by a single participant are assigned one alignment score. Comments separated by ‘//’ have a different stance focus and are coded separately.

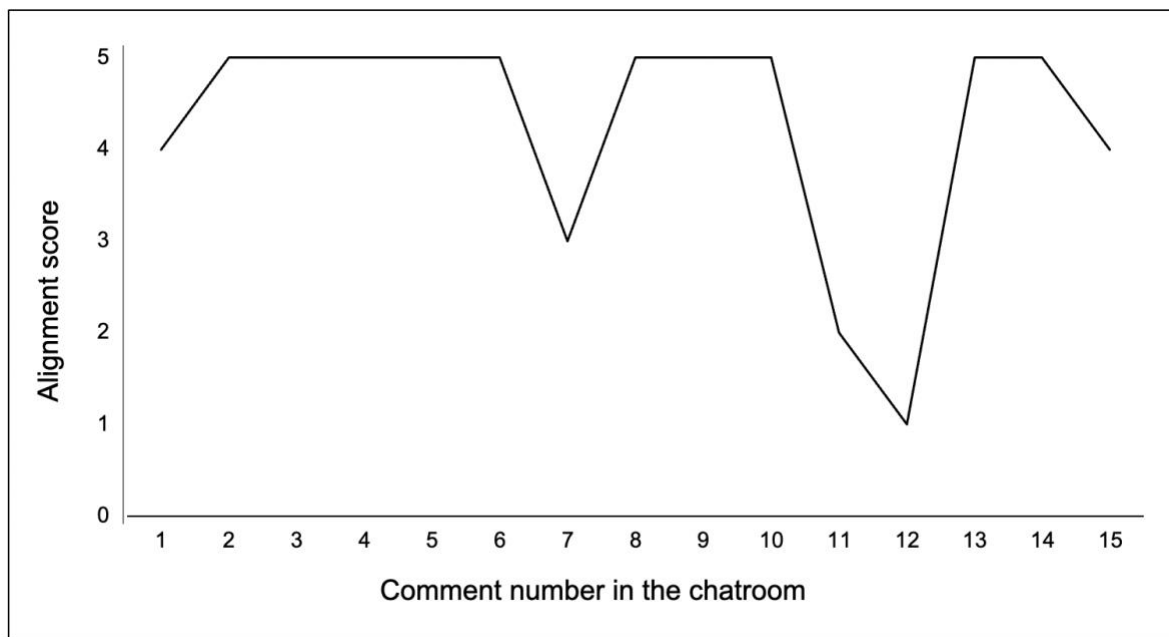


Figure 9. Alignment score for each comment in relation to prior speaker in Gold team discussion (1 = disaligned, 5 = aligned)

Let us start with the speech of the two defectors. In utterance 5, Gold1F structurally aligns with Gold2M’s proposed collective strategy early on, resonating with his syntactic construction (Du Bois and Giora 2014) and upgrading the collective benefit of altruism from “you” to an inclusive “us”. This draws attention to her trustworthiness without an explicit promise, and she does not participate further in the discussion. As Goffman (1956: 41) notes, “communication techniques such as innuendo, strategic ambiguity, and crucial omissions allow the misinformer to profit from lies, without, technically, telling any.” Gold1F uses vagueness to avoid lying (much like Orange5F did), while other defectors in the study were willing to manipulate their group members, as we will see with Gold5M. The upshot of Gold1F’s comment is a *perceived* consensus on the part of Gold2M (cf. Bouas and Komorita 1996). In Round 2, when Gold1F’s defection was revealed in the Online Chat, he expressed his anger through Multicultural London English (MLE) (c.f. Cheshire et al. 2011): “GOLD1F is *moving mad!*” before arguing that “Everyone agreed to share 40 to increase the pot - more squeezes shared more money earned. someone is obviously content being selfish and earning less, than sharing and earning more. (...) Que pasa!” Interestingly, at this later stage, Gold1F responds with entirely explicit reasoning, and with none of her earlier structural alignment now apparent: “Once you have calculated the maths, it allows you to make the right contribution.”

The other defector, Gold5M, has proven himself in Round 1 by contributing 30 squeezes to the group. Like Gold1F, he triggers an “activation of affinities across utterances” (Du Bois and Giora 2014: 356) in utterances 8, 10, and 13 to emphasize the precise financial reward of cooperating. His strong epistemic forms (“exactly”, imperatives, direct questioning) support an authoritative position within the group, as we saw in both previous teams; in this case it also avoids arousing suspicion. In Round 2, Gold5M’s response to the punishments from his disgruntled group shows the explicit insincerity of his utterances in Round 1: “It benefits me when I convince everyone to contribute as much as possible while I contribute none. This allows me to get the most. Sorry guys.”

In addition to these sincerity and non-participation issues, a further source of dissonance in the exchange is an unsettled mix of registers and frames. Gold2M's use of the slang term "deffo" ('definitely') and an omitted pronoun and article conveys informality, as does his ally Gold5M's "yea" later. By contrast, Gold4M enters the exchange with a more formal, business voice. The term "profitable", the only use of the word in all 18 discussions, and whose root ("profit") features in the top 100 keywords of Nelson's (2000) Business English Corpus (BEC). We see an influence of business speak in later comments by Gold2M and Gold5M, e.g., "maximize", "reward", "contribution", "approve", "best possible solution" (once again, "solution" was found nowhere else in the discussions but featured in Nelson's (2000) list of positive keywords in the BEC). Gold4M's comments are also marked by weakeners "quite" and "we can" and a lack of affective display, all of which introduce ambiguous personal commitment. Gold4M may benefit from this discursive distancing when he later makes a counteroffer to Gold5M's proposal.

Gold2M's enthusiastic endorsement of Gold5M forms an alliance; unfortunately for Gold2M, defectors can interpret pledges as a sign that the cooperator is an easy target (Camera et al. 2013). Gold5M conceals this well with his high degree of alignment in comment 10. The final segment involves a negotiation between Gold2M/Gold5M and Gold4M, who is eventually convinced to cooperate as indicated by his agreement the following afternoon.

Once the results are revealed in Round 2, Gold2M's sense of betrayal is evident. He not only scorns Gold1F, as we saw earlier, but also Gold5M: "GOLD5M is moving loose!" His strong negative alignment is paralleled by a structural shift away from his brief use of the business register to MLE. Gold3F's response confirms the sense that group members produced *and* perceived the discussion as formulaic of a contractual business meeting: "Well done to those who kept to the agreement. Clearly some group members cannot help being out for themselves!!" Gold4M focused solely on the moral contract: "Greed can be seen from 2 guys here." We end with this example to acknowledge that talk can be cheap. All the same, Gold1F avoided explicit deception, which suggests a strong constraining role of reputation. Combined with the binding promises of the remaining three players and their intense negative reaction to Gold3M's broken promise, talk is widely used even in this group as a sincere signal of cooperative intentions. Finally, it is important to bear in mind that the groups showed a fair level of correspondence between talk and behaviour overall (Table 2).

Other teams

How representative are these case studies of the teams as a whole? Teams with low consensus scores in Table 2 were negatively impacted by a combination of issues, starting with weak leadership, which we have seen affected the Orange team. Among the two most unsuccessful groups, both of which contributed *fewer* squeezes in Round 2, discussions were dominated by a single participant who then brought personal and group competency problems to the fore. In the Turquoise team, only one participant increased her contribution in Round 2, a non-native female (Turquoise2F) who participated heavily at the beginning of the discussion. Her leading role in conversation and rather formal and stilted language resembled that of Orange3F. She disaligned from the group with doubts over their altruism ("Also, at the very beginning I was a bit skeptical on how much to share as I wanted to see how the rest of the group would behave"), and this mistrust is cited later by a defector in Round 2 ("Oh wow...well done Turquoise2F! I didn't give more squeezes because I thought everyone was going to pull a 0 for the group pot in this round.") Similarly, in the Scarlet team, 12 out of 18 comments were from a female participant who was simultaneously figuring out the task and trying to guide the group strategy. The leadership of these two participants contrasts dramatically with Green1M's adept and adaptive leadership style.

Other unsuccessful groups encountered related problems of distrust, low expectations of cooperation (e.g., expressing surprise when presented with evidence of cooperation), unfamiliarity with the format of chatroom discussion (e.g., awkwardly repeating previous posts), and misaligned comments. These factors negatively impacted the consensus process in two ways: the quantity and quality of participants' promises to cooperate. This is supported by the fact that in the bottom four groups, there were not only far fewer promises, but these were also more vague in terms of commitment (e.g., a promise to contribute more rather than all 40). A non-committal promise or no promise at all is a safe option when distrust is present, leadership is shaky, or understanding is low, and so these triggers of disalignment led to an inability to reach a consensus to cooperate.

By contrast, the top four groups had a far higher number of promises, and more that specified an exact contribution of 40 squeezes. Did this strong consensus arise in similar ways to our case study of the Green team above? The most successful group, Mauve, had an initial misunderstanding that was resolved, after which all participants were highly aligned, making five promises to contribute all 40 squeezes in Round 2. More importantly, several used a gameshow register just like the Green team, with comments such as "TEAM MAUVE!!!" and "This reminds me of the show Golden Balls" (to which someone replied: "I'm more of a Crystal Maze man myself.") Cooperating teams showed further similarities to those discussed here, including socialist discourse, resolution of misunderstanding, and playfully stern exhortations to cooperate that received widespread uptake.

Conclusion

Our analysis shows that numerous elements are at play during online chat communication, and the relative success an actor has in imposing a given strategy depends on uptake within the group. For example, one successful group may fall in line with a dominant but persuasive leader whose opinion serves as a reference point, while another may converge upon an egalitarian norm for stance convergence. A mixed methods approach examining these qualitative underpinnings of a quantitative outcome helps to harness the potential of natural language while operationalising it as a measurable factor in behaviour. Specifically, our work builds on previous research in two ways. First, our quantitative consensus analysis confirms that different types of communication affect cooperation (as per McClung et al. 2017; Goren and Bornstein 2000; Kagel 2018), but it goes further by identifying important differences between the categories, and assigning each one a numerical value to explore its role in supporting a consensus to cooperate. In doing so, we show how linguistic subtleties relate to varying degrees of cooperative behaviour. Second, unlike previous work, our alignment analysis captures the intersubjective nature of social dilemmas by studying the content and structure of the whole interaction, while our register analysis highlights how the myriad social voices that we use in everyday communication help and hinder cooperative efforts.

Despite the non-deterministic nature of cooperation through talk, our study indicates that certain registers are regularly associated with cooperative behaviour: gameshow, familial, and socialist styles of interaction. These voices were used to align with or disalign from others in three crucial ways: (1) foster group identity through encouragement and clarification of misunderstandings; (2) instil a moral norm of cooperation by expressing high expectations or delivering punishment; and/or (3) ensure players publicly commit to cooperate via strong leadership. Even beyond our case studies, successful groups displayed these three strategies to varying degrees (Yellow, Rose, Mauve), advancing findings of past proposals (moral norms, Bicchieri and Lev-On 2007; public commitment, Kerr and Kaufman-Gilliland 1994, and group identity, Orbell et al. 1988). The cases here show how many of these isolated factors work together in natural spoken language.

On the other hand, political and business registers featured in less cooperative groups. The individualistic agenda of figures in the world of politics and business may not only evoke associations of mistrust, sales pitches, and heavy-handed leadership, but they may also reduce obligation due to the high social distance of such voices. These effects were observable in the three case studies as well as in other unsuccessful groups, along with low expectations (Black, Silver) and unresolved confusion (Scarlet).

We note some inevitable limitations of this study. The timeframe between the first Online Chat and Round 2 was not strictly identical. On average this was 2-6 days, but it varied by group, which may have meant that the impact of the Online Chat on Round 2 behaviour was greater among groups who completed Round 2 sooner compared to those who completed it later. For example, a 'miffed' participant in the Blue team admitted to the experimenter that she was initially very upset by her group's behaviour. Balliet's (2010) meta-analysis suggests that the impact of communication on cooperation is not different depending on whether it is before or during the game, so these differences may not be too consequential. Nevertheless, to mitigate this variation, every participant was shown the Online Chat before completing Round 2 to ensure equal exposure and therefore reduce differences in emotional response to the discussion. Asking participants about the extent to which they thought the Online Chat impacted their behaviour and testing for correlations with the timeframe would be a useful empirical check in this regard. More generally, while studies have focused on FtF communication during the decision-making process, it is also important to study CMC that occurs before the decision-making process. From an organisational perspective, channels such as Slack or the instant messenger function in MS Teams are increasingly popular due to pressures of time and space (Berry 2006), and more so since a hybrid model of working emerged in March 2020. Crucially, these new communication channels are revolutionising the way we cooperate (Conrads and Reggiani 2014), such that the communicative event takes place neither in-person nor concurrently with the decision-making process. This area also offers a fruitful avenue for future research. The presence of non-verbal cues in video-calls, for example, would no doubt add nuance to expressions of register and alignment between speakers, but would this help or hinder cooperation?

A case study approach also faces problems of generalisability and subjectivity. We avoid claims of direct causation here, given the nature of the data, but we offer the statistically significant consensus effect from the wider study (Table 2) as one way to link the moment-to-moment negotiation of talk to behavioural outcomes. Understanding how talk builds consensus may be challenging in methodological terms but the richness of this data justifies its integration into the scientific study of collective action. Equally, current quantitative methods studies do not attend to the details of communication, but they are valuable in identifying aggregate effects. We have therefore proposed both quantitative (correlation of aggregate consensus score with subsequent behaviour; degrees of inter-personal alignment in real-time) and qualitative (choice of register and imposition of behavioural norms via frames) tools to better understand the relationship between communication and behaviour.

We cannot rule out the role of individual differences driving the above dynamics. There has been increased interest in explaining behavioural heterogeneity through the HEXACO personality framework (Ashton and Kibeom 2007). Of particular relevance to social dilemmas is the Honesty-Humility (HH) scale, which includes greed avoidance, sincerity, fairness and modesty (Lee et al. 2008), and research suggests that fairness and greed avoidance are strongly related to prosocial behaviour (Hilbig et al. 2014). Future work exploring verbal interaction among people with similar levels of HH would allow us to tease apart the role of talk and individual differences in cooperation.

In sum, social dilemmas affect society at all levels, from international agreements to intimate relationships. We have argued that enriching traditional studies on cooperation requires extending the line of inquiry to verbal interaction (Goffman 1961; Sally 2002, 2005; Vollmer 2013). By exploring the registers available to individual actors and the alignments between actors, we hope to have highlighted the potential for fruitful interdisciplinary dialogue between sociology, economics, and linguistics.

References

- Agha, A (2005) Voice, footing and enregisterment. *Journal of Linguistic Anthropology* 15(1): 38-59.
- Archakis A, Tsakona, V (2010) ‘The wolf wakes up inside them, grows werewolf hair and reveals all their bullying’: The representation of parliamentary discourse in Greek newspapers. *Journal of Pragmatics* 42: 912-923.
- Ashton MC, Kibeom L (2007) Empirical, theoretical, and practical advantages of the HEXACO Model of Personality Structure. *Personality and Social Psychology Review* 11(2): 150-166.
- Bates, D, Maechler, M, Bolker, BM, Walker, SC (2015) Fitting linear mixed- models using lme4. *Journal of Statistical Software* 67(1): 1-48.
- Balliet, D (2009) Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution* 54(1): 39-57.f
- Berry, GR (2006) Can computer-mediated asynchronous communication improve team processes and decision-making? *Journal of Business Communication* 43(4): 344-366.
- Bicchieri, C (2002) Covenants without swords: Group identity, norms, and communication in social dilemmas. *Rationality and Society* 14(2): 192-228:
- Bicchieri, C, Lev-On, A (2007) Computer-mediated communication and cooperation in social dilemmas: An experimental analysis. *Politics, Philosophy and Economics* 6(2): 139-168.
- Bouas, KS, Komorita, SS (1996) Group discussion and cooperation in social dilemmas. *Personality and Social Psychology Bulletin* 22(11): 1144-1150.
- Brosig, J, Ockenfels, A, Weimann, J (2001) The effect of communication on cooperation. *German Economic Review* 4: 217-241.
- Brown, P, Levinson, SC (1987) *Politeness: Some Universals in Language Usage*. Cambridge, United Kingdom: Cambridge University Press.
- Burnett, H (2019) Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy* 42: 419-450.
- Camera, G, Casari, M, Bigoni, M (2013) Binding promises and cooperation among strangers. *Economic Letters* 118: 459-461:

- Cheshire, J, Kerswill P, Fox, S, Torgersen, E (2011) Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics* 15(2): 151-196.
- Conrads, J, Reggiani, T (2016) The effect of communication channels on promise-making and promise-keeping: Experimental evidence. *Journal of Economic Interaction and Coordination* 12(3): 595: 611.
- Coreen, F, Sandler, S (2014) Polyphony, ventriloquism, and constitution: In dialogue with Bakhtin. *Communication Theory* 24(3): 225-244.
- Du Bois, J (2007) The stance triangle. In: Engelbreston, R (ed) *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*. Amsterdam: John Benjamins, pp. 139–182.
- Du Bois, J (2011) Co-opting intersubjectivity: Dialogic rhetoric of the self. In: Meyer, C, Girke, F (eds) *The Rhetorical Emergence of Culture*. Oxford: Berghahn, pp. 52-83.
- Du Bois, J, Giora, R (2014) From cognitive-functional linguistics to dialogic syntax. *Cognitive Linguistics* 25(3): 351-357
- Du Bois, J, Kärkkäinen, E (2012) Taking a stance on emotion: Affect, sequence, and intersubjectivity in dialogic interaction. *Text and Talk* 32(4): 433-451.
- Duranti, A (2006) Narrating the political self in a campaign for U.S. Congress. *Language in Society* 35(4): 467-497.
- Dunning, D, Anderson JE, Schlösser T, Ehlebracht, D, Fetchenhauer, D (2014) Trust at zero acquaintance: More a matter of respect than expectation of reward. *Journal of Personality and Social Psychology* 107(1): 122-141.
- Engelmann, JB, Fehr, E (2017) The neurobiology of trust: The important role of emotions. In: Van Lange, P, Rockenbach, B, Yamagishi, T (eds) *Social dilemmas: New Perspectives on Reward and Punishment*. New York: Oxford University Press, pp. 33-56).
- Fezter, A (2014) I think, I mean, I believe in political discourse. Collocates, functions and distribution. *Functions of Language* 21(1): 67-94.
- Fetzer, A, Bull, P (2012) Doing leadership in political speech: Semantic processes and pragmatic inferences. *Discourse and Society* 23(2): 127-144
- Frolich, N, Oppenheimer, J (1998) Some consequences of email vs face-to-face communication in experiment. *Journal of Economic Behavior and Organization* 35: 389-403.
- Gächter, S, Fehr, E (1999) Collective action as a social exchange. *Journal of Economic Behavior and Organization* 39(4): 341-369.
- Goffman, E (1956) *The Presentation of Self in Everyday Life*. Edinburgh: University of Edinburgh.

- Goffman, E (1961) *Encounters: Two Studies in the Sociology of Interaction*. Indianapolis: Bobbs-Merrill.
- Goffman, E (1974) *Frame Analysis: An Essay on the Organization of Experience*. Cambridge, MA: Harvard University Press.
- Goren, H, Bornstein, G (2000) The effects of intragroup communication on intergroup cooperation in the repeated intergroup prisoner's dilemma (IPD) game. *Journal of Conflict Resolution* 44(5): 700- 719.
- Halvorsen, K (2015) Questions as interaction resource in team decision making. *International Journal of Business and Communication* 55(1): 69-93.
- Hilbig, BE, Glockner, A, Zettler, I (2014) Personality and prosocial behaviour: Linking basic traits and social value orientations. *Journal of Personality and Social Psychology* 107(3): 529-539.
- Houghton, KJ, Upadhyay, SS, Klin, CM (2018) Punctuation in text messages may convey abruptness. *Computers in Human Behavior* 80: 112-121.
- Jacknick, CM, Avni, S (2017) Shalom, bitches: Epistemic stance and identity work in an anonymous online forum. *Discourse, Context & Media* 15: 54-64.
- Jäger, G (2008) Applications of game theory in linguistics. *Language and Linguistics Compass* 2(3): 406-421.
- Jaffe, A (2009). The sociolinguistics of stance. In: Jaffe, A (ed) *Stance: Sociolinguistics Perspectives*. Oxford: Oxford University Press, pp. 3-28.
- Joseph, J (2006) *Language and Politics*. Edinburgh: Edinburgh University Press.
- Kagel, JH (2018) Cooperation through communication: Teams and individuals in finitely repeated Prisoners' dilemma games. *Journal of Economic Behavior and Organization* 146: 55-64.
- Kärkkäinen, E (2006) Stancetaking in conversation: From subjectivity to intersubjectivity. *Text and Talk* 26(6): 699-731.
- Kiesling, SF, Pavalanathan, U, Fitzpatrick, J, Han, X, Eisenstein, J (2017) Interactional stancetaking in online forums. *Computational Linguistics* 44(4): 683-718
- Kerr, NL, Kaufman-Gilliland, CM (1994) Communication, commitment, and cooperation in social dilemma. *Journal of Personality and Social Psychology* 66(3): 513- 529.
- Kydd, A (2000) Overcoming mistrust. *Rationality and Society* 12(4): 397-424.
- Lee, K, Ashton, MC, Morrison, DL, Cordery, J, Dunlop, PD (2008) Predicting integrity with the HEXACO Personality Model: Use of self- and observer reports. *Journal of Occupational and Organizational Psychology* 81(1): 147–167.

- Lenger, A (2019) The rejection of qualitative research methods in economics. *Journal of Economic Issues* 53(4): 946-965.
- Lev-On, A, Chavez, A, Bicchieri, C. (2010). Group and dyadic communication in trust games. *Rationality and Society* 22(1): 37-54.
- Ludwiczak, A, Adams, Z, Osman, M. Submitted. Actions do not always speak louder than words: the role of social incentives when choosing and executing cooperative effort.
- Majeski, SJ, Fricks, S (1995) Conflict and cooperation in international relations. *Journal of Conflict Resolution* 39(4): 622-645.
- McClung, JS, Placi, S, Bangerter, A, Clement, F, Bshary, R (2017) The language of cooperation: Shared intentionality drives variation in helping as a function of group membership. *Proceedings of the Royal Society B. Biological Sciences*, 284: 20171682-20171689.
- Milkovich, M, Sitarica, A (2017) Epistemic modality in political discourse. *Research Result, Social Studies and Humanities* 3(1): 75-79.
- Muehlbacher, S, Kirchler, E (2009) Origin of endowment in public good games: The impact of effort on contributions. *Journal of Neuroscience, Psychology, and Economics* 2(1): 59-67.
- Myers-Scotton, C, Bolonyai, A (2001) Calculating speakers: Codeswitching in a rational choice model. *Language in Society* 30(1): 1-28.
- Nelson, M (2000) Corpus-based study of the lexis of Business English and Business English Teaching Material. [Doctoral dissertation, University of Manchester]
<http://users.utu.fi/micnel/thesis.html>
- Ochs, E, Taylor, C (1995) The “Father knows best” dynamic in dinnertime narratives. In: Hall, K, Bucholtz, M (eds) *Gender Articulated: Language and the Socially Constructed Self*. London: Routledge, pp. 97-120.
- Orbell, JM, Van de Kragt, AJ, Dawes, RM (1988) Explaining discussion-induced cooperation. *Journal of Personality and Social Psychology* 54(5): 811-819.
- Perry, I (2004) *Prophets of the Hood: Politics and Poetics in Hip Hop*. Durham and London: Duke University Press.
- Rice, RE (1984) *The New Media: Communication, Research and Technology*. Beverley Hills: Sage.
- Roe, K, Messing, V, Vandebosch, H, Van den Bulck, J (1996) Television game show viewers: A cultivated audience? *Communications* 21(1): 49-64.
- Sally, D (1995) Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958-1992. *Rationality and Society* 7(1) 58-92.

- Sally, D (2002) ‘What an ugly baby!’ Risk dominance, sympathy, and the coordination of meaning. *Rationality and Society* 14(1): 78-108.
- Sally, D (2005) Can I say “bobobo” and mean “There’s no such thing as cheap talk”? *Journal of Economic Behavior and Organization* 57: 245-266.
- Samuelson, CD, Watrous-Rodriguez, KM (2009) Group discussion and cooperation in social dilemmas: Does the medium matter? In: Kramer, RM, Tenbrunsel, AE, Bazerman, MH (eds) *Social Decision Making: Social Dilemmas, Social Values, and Ethical Judgments*. New York, NY: Routledge, pp.13-46.
- Tajfel, H, Turner, JC (1979) The social identity theory of intergroup behavior. In: Austin, WG, Worchel, S (eds) *The Social Psychology of Intergroup Relations*. Chicago: Nelson-Hall, pp. 7-24.
- Torsvik, G, Molander, A, Tjøtta, S, Kobbeltvedt, T (2011) Anticipated discussion and cooperation in a social dilemma. *Rationality and Society* 23(2): 199-216.
- Twenge, JM, Campbell, WK, Gentile, B (2012) Increases in individualistic words and phrases in American books, 1960-2008. *PLoS ONE*, 7(7).
- Vandergriff, I (2013) Emotive communication online: A contextual analysis of computer-mediated communication (CMC) cues. *Journal of Pragmatics* 51. 1-12.
- Van der Valk, I (2003) Right-Wing parliamentary discourse on immigration in France. *Discourse & Society* 14 (3): 309-348.
- Van Leeuwen, T (2007) Legitimation in discourse and communication. *Discourse and Communication* 1(1): 91-112.
- Vollmer, H (2013) What kind of game is everyday interaction? *Rationality and Society* 25(3): 370-404.
- Vukovic, M (2014). Strong epistemic modality in parliamentary discourse. *Open Linguistics* 1: 37-52.