The Effect of Collaborative Triadic Conversations in Noise on Decision-Making in a General-Knowledge Task

Ingvi Örnolfsson[,1] (iD), **Axel Ahrens**[1] (iD), **Torsten Dau**[1] (iD), **and Tobias May**[1] (iD)

Abstract

Collaboration is a key element of many communicative interactions. Analyzing the effect of collaborative interaction on subsequent decision-making tasks offers the potential to quantitatively evaluate criteria that are indicative of successful communication. While many studies have explored how collaboration aids decision-making, little is known about how communicative barriers, such as loud background noise or hearing impairment, affect this process. This study investigated how collaborative triadic conversations held in different background noise levels affected the decision-making of individual group members in a subsequent individual task. Thirty normal-hearing participants were recruited and organized into triads. First, each participant answered a series of binary general knowledge questions and provided a confidence rating along with each response. The questions were then discussed in triads in either loud (78 dB) or soft (48 dB) background noise. Participants then answered the same questions individually again. Three decision-making measures – stay/switch behavior, decision convergence, and voting strategy - were used to assess if and how participants adjusted their initial decisions after the conversations. The results revealed an interaction between initial confidence rating and noise level: participants were more likely to modify their decisions towards high-confidence prior decisions, and this effect was more pronounced when the conversations had taken place in loud noise. We speculate that this may be because low-confidence opinions are less likely to be voiced in noisy environments compared to high-confidence opinions. The findings demonstrate that decision-making tasks can be designed for conversation studies with groups of more than two participants, and that such tasks can be used to explore how communicative barriers impact subsequent decision-making of individual group members.

Introduction

The act of speech communication is often framed using the source-message-channel-receiver model of communication, also known as the linear model of communication (Berlo, 1960). While effective communication relies heavily on receiving and interpreting auditory signals, it is essential to recognize that communication is fundamentally interactive. Various experimental paradigms have empirically demonstrated this (Bavelas et al., 2000; Fay, Garrod & Carletta, 2000; Garrod & Pickering, 2009; Schober & Clark, 1989). In their critique of analyzing conversation solely through turn-taking behavior, O'Connell et al. (1990) emphasized that 'listeners are essential and active parties of conversations'. Thus, evaluating communication difficulty and success must go beyond hearing, listening, and comprehension to include interpersonal interaction (Carlile & Keidser, 2020; Kiessling et al., 2003).

In recent years, there has been growing interest in studying conversational behavior in realistic, face-to-face interactions affected by communicative barriers such as noise and hearing impairment (Buchholz et al., 2022; Hadley et al., 2019; Hadley et al., 2021; Miles et al., 2023; Petersen, 2024; Petersen et al., 2022). These studies have proposed various ways to conceptualize and predict communication

[1]Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark

Corresponding Author:
Ingvi Örnolfsson, Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark.
Email: rinor@dtu.dk

Data Availability Statement included at the end of the article

difficulty. Beechey et al. (2018) used subjective effort ratings to assess dyadic communication difficulty in different noise levels, revealing that various speech production measures can be used to predict communication difficulty. Miles et al. (2023) used communicative breakdowns as indicative of communication difficulty and showed that dyads use synergistic strategies such as postural movement coordination and speech level adaptations when the background noise level increases. While turn-taking measures (e.g., floor-transfer offsets, speaking time and utterance durations) have also been suggested to be related to communication difficulty (Sørensen et al., 2021), it remains unclear how strong this relationship is (Petersen, 2024; Watson et al., 2019).

The wide diversity of methods for measuring communicative difficulties suggests that defining quantitative success criteria in interactive communication studies is challenging. One reason may be that the objectives of communicative interaction are often not very precisely defined, particularly in the non-task-oriented conversations common in hearing research (Hadley & Ward, 2021; Hadley et al., 2021; Miles et al., 2023; Petersen, 2024). O'Connell et al. (1990) argue that the goals of conversations vary widely depending on context. Conversations can, for example, be rooted in empathy, deception, encouragement, or blame; they can be collaborative or antagonistic; and they can be oriented towards a tangible outcome, obtaining mutual understanding, or simply relaxation and socializing. No single experimental paradigm can encompass all these goals. Thus, defining a clear objective for conversations used in research may be useful when studying communication difficulty and success.

One way to define such an objective is by using tasks that require participants to collaborate verbally to achieve a predefined goal, with objective outcomes indicating successful collaboration. Examples of such tasks include the *tangram maze task*, where participants navigate a maze whose layout is divided between them (Beechey et al., 2018, 2019), and the *diapix task* (Baker & Hazan, 2011; Petersen et al., 2022; Sørensen et al., 2021; Watson et al., 2019), a spot-the-difference task where each participant sees only one version of the image. These tasks rely on joint decisions, with correct decisions reflecting successful information exchange (Beechey et al., 2019). A recent study further supported this approach, identifying information exchange as a key indicator of communicative success in both normal-hearing and hearing-impaired individuals (Nicoras et al., 2022).

A limitation of the tangram and diapix tasks is that they are designed for only two participants, rendering them impractical for studies involving larger groups where collaborative decision-making outcomes are desired. While two-person conversations are the most frequent single group size in everyday life, many conversations involve more than two interlocutors (Peperkoorn et al., 2020), a scenario which is particularly challenging for hearing-impaired individuals (Kiessling et al., 2003; Nicoras et al., 2022). To date, studies involving more than two participants have mostly relied on task-free dialogue (e.g.,

Hadley & Ward, 2021; Hadley et al., 2021; Lu et al., 2021; Petersen, 2024). Developing tasks for larger groups could be useful for understanding the negative consequences of communicative barriers such as noise and hearing loss on collaborative decision-making.

In this study, we introduce a new communication task based on decision-making that accommodates any group size. The task is built on a framework commonly used in the literature on decision-making and social influence in groups and collaborative settings (Bahrami et al., 2010; Bang et al., 2014; Fay et al., 2000; Keshmirian et al., 2022; Koriat, 2012; Mahmoodi et al., 2018). A typical structure in these studies, which we emulated here, involves participants first making individual decisions about a certain question (related to, for example, a perceptual task, an ethical dilemma, or a preference judgment), followed by a discussion round. After the discussion, a posterior decision is made on the same question, either jointly in the group or individually. In our implementation of the task, participants first answered general-knowledge questions individually, then discussed them in triadic groups before answering the same questions individually again. This task differs from the diapix and tangram maze tasks in several ways, most notably by introducing an element of epistemic uncertainty and by using individual rather than joint decision-making. Using this task, we investigated how individual participants' decisions changed after triadic discussions in both an easy and a difficult communication scenario. We varied communication difficulty by presenting multi-talker background noise at two different intensities. Three decision-making measures were used to analyze how individual participants' decisions changed after the conversations; one was drawn from the literature on group decision-making, and two new measures were devised for this study.

Methods

Participants

The study comprised 30 participants organized into groups of ten triads. All participants, aged between 20 and 35 years, were native Danish speakers with self-reported normal-hearing status. Except for two pairs, participants were unfamiliar with each other prior to the experiment. The experiment was conducted in Danish and took approximately 2.5 h including participant instruction, break, and debriefing. Time spent in the experiment was approximately 1.5 h. Informal interviews with participants of pilot studies suggested that this duration was reasonable and not too mentally taxing. Participants received compensation for their time after providing informed consent. Ethical approval was obtained from the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

Environment and Experimental Setup

The participants were seated in an equilateral triangle configuration, facing the other two group members, as illustrated in Figure 1b. The distance between participants was

approximately 1.5 m. They wore eye-tracking glasses capturing point-of-view footage, eye-gaze data, and pupil dilation. Additionally, three microphones were utilized, including a pair of in-ear binaural microphones and a cheek-mounted microphone; however, data from these devices were not analyzed in this study.

The group was surrounded by eight loudspeakers (Dynaudio BM6P), arranged in a ring with a radius of 2.4 m. To minimize visual distractions, a circular black curtain fully enclosed the participants and the loudspeakers. Each loudspeaker played a separate Danish monologue (Ahrens & Lund, 2022), resulting in a spatially distributed multi-talker masker. The monologues lasted approximately 90 s each and were looped for the duration of the conversation. The loudspeakers were driven by a sonible d:24 amplifier. The masker was presented at sound pressure levels (SPLs) of either 48 dB or 78 dB, corresponding to the soft and the loud conditions, respectively. The simultaneous presentation of multiple masking speech sources rendered them individually unintelligible in both conditions. The two noise levels were selected based on previous studies of dyadic and triadic interactions in noise, which have shown that 78 dB is sufficient to elicit behavioral changes and prompt participants to report modifying their communication strategies (Beechey et al., 2019; Hadley et al., 2021; Miles et al., 2023; Petersen, 2024).

Task

The initial task for the participants involved responding individually to a series of binary general-knowledge questions categorized into three topics: Hollywood movies (identifying the oldest of two movies), Copenhagen landmarks (determining which of two locations is closest to the city center), and European countries (determining which of two countries has the most inhabitants). Each topic comprised two lists of 28 questions, one for each acoustic noise condition. Consequently, each list contained 28 trials, formulated by employing all unordered pairs from the eight items associated with that topic (e.g., eight Hollywood movies). Before the primary experiment, the group underwent a brief trial round on a different topic not used in the study. This allowed participants to familiarize themselves with the task, the technical interface, and with each other, thus overcoming any initial awkwardness in their conversations.

Questions were presented on a touch-screen tablet, showing a visual illustration of the two options along with accompanying labels. The participants were instructed to select an option and provide a confidence rating, expressed as a percentage between 50% and 100%, with 50% indicating no preference for either option, while 100% meant absolute certainty in the decision (Figure 1a). They were asked to interpret the scale as indicating their estimated probability of having answered the question correctly, i.e., a metacognitive judgment. In the analysis stage, this rating was assumed to be related to the confidence with which participants would express their opinion. While this assumption was not tested directly in the present study, previous studies on joint decision-making suggest that this is a reasonable assumption, at least for dyads performing joint perceptual tasks (Bahrami et al., 2010; Fusaroli et al., 2012).

After the initial set of 28 questions, a conversation round followed, during which the participants discussed their answers with the other group members. Participants had previously been told that they would have to repeat the questions individually after the conversation round; the aim of the conversation round was for the participants to share any information they
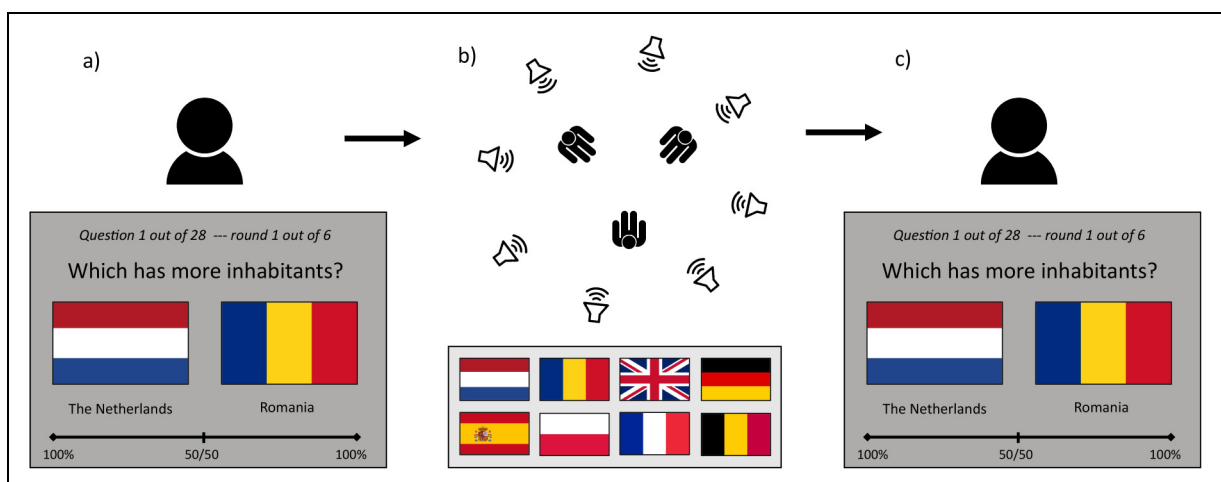

Figure 1. Overview of the experimental procedure. a) Participants initially made individual decisions on a series of binary general knowledge questions, submitting decisions along with a confidence rating using a continuous scale. b) Subsequently, participants engaged in group conversations with two other members, with the aim of improving each other's answers. The conversation took place in either loud background noise (78 dB spatialized 8-talker babble) or in soft noise (48 dB spatialized 8-talker babble). c) Following the conversation, participants independently and privately repeated the same questions.

held that might be important for answering the questions. Participants were instructed to engage with the task collaboratively, emphasizing the importance of improving the subsequent performance of all group members, not just oneself. To facilitate the discussion, each participant was given a sheet displaying the eight items from the preceding question round during the conversation (see Figure 1b). During the conversations, no structure was enforced upon participants; they were free to ask each other questions, engage in discussions of particular questions they recalled, or discuss the items more generally. A very common strategy that groups ended up employing was a ranking approach, where, at or near the end of the discussion round, they would attempt to go over the item list and rank it according to the criteria that the questions were concerned with. Once a 10-min time limit was reached or the conversation concluded naturally, participants completed a short questionnaire with five questions related to 1) how effortful the conversation was, 2) how easy it was to understand other group members, 3) how easy it was to express themselves, 4) how engaged they felt in the conversation, and 5) how they perceived the flow of the conversation. Each question was answered using a discrete 11-point scale. After completing the questionnaire, participants individually answered the 28 questions from the pre-conversation round again (see Figure 1c). At the end of each post-conversation round, participants received a score based on how well they, as a group, scored on the 28 questions in the post-conversation round. This score was calculated as the mean percent correct across the three participants' post-conversation decisions. This entire process (28 pre-conversation questions – conversation round – questionnaire – 28 post-conversation questions) was repeated six times, once for each of the three topics and in each of the two noise conditions. The order of topics and conditions was randomized between groups, with the additional constraint that the same topic could never appear twice in a row. A brief break was incorporated after the third question round.

Decision-Making Measures

To evaluate how conversations affected individual participants' post-conversation decisions, we analyzed the pre- and post-conversation decisions of group members using three different measures: 1) individual stay/switch behavior, 2) group and pair convergence within the group, and 3) the group's voting strategy. The first two measures were designed specifically for this study, while the third is an existing measure from the literature on group decision-making (Koriat, 2012; Meyen et al., 2021).

The terms *staying* and *switching* refer to the post-conversation decisions of individuals when another group member *disagrees* with their initial decision. Disagreement, in this sense, is determined entirely by the initial decisions of participants; it does not imply that disagreement was necessarily verbalized during the conversation. If two participants each chose a different option in a given pre-conversation trial, this

will be referred to as a *disagreement trial*. For each disagreement trial, the post-conversation decisions of disagreeing members were classified into one of two behaviors; they either *stay* with their initial decision or *switch* to the alternative decision previously chosen by the other member. Each of the three distinct pairs of individuals within a group were analyzed separately, disregarding the prior decision of the third member. As an example, consider a case where the prior decisions of a group are [A A B], and the posterior decisions are [A A A], with A and B denoting the two options (e.g., "The Netherlands" and "Romania" in the example shown in Figure 1). As there is no prior disagreement between members one and two, this pair is ignored. Analyzing the pair consisting of members one and three reveals that member one stayed ($[A \rightarrow A]$), while member three switched ($[B \rightarrow A]$). Similarly, in the pair of members two and three, member two stayed ($[A \rightarrow A]$) while member three switched ($[B \rightarrow A]$).

The second measure, *convergence,* was defined in two distinct ways: at the pairwise level and at the group level. Pairwise convergence was defined as trials where a pair transitions from disagreeing before the conversation to agreeing afterwards, e.g., $[A B] \rightarrow [A A]$. Occasionally, pairs might "convince each other", maintaining disagreement but adopting each other's prior decisions, e.g., $[A B] \rightarrow [B A]$. This behavior was not considered as pairwise convergence. Group convergence was defined as trials where both options were present in prior decisions, and all group members chose the same option after the conversation, e.g., $[B A B] \rightarrow [A A A]$.

The third measure examined individuals' voting strategies. Specifically, two existing models of decision-making in groups - *majority voting* and *confidence slating* - were compared. Majority voting reflects the tendency of posterior decisions to follow the most popular prior decision, while confidence slating reflects the tendency of posterior decisions to follow the prior decision favored by the most confident individual. While these models typically relate to groups making joint decisions, they are straightforward to apply to the present case where posterior decisions are made individually. For example, if a groups' prior decisions are [A **B** A], with boldface indicating the most confident member, any member selecting A after the conversation will be classified as an instance of majority voting. Conversely, choosing B would be categorized as confidence slating. Notably, confidence slating and majority voting are not mutually exclusive, as the most confident member can be part of a majority. In order to make the distinction between these two strategies as clear as possible, the voting strategy analysis was limited to trials where the most confident member was in the minority.

Data Preprocessing and Statistical Analyses

The participants' confidence ratings were transformed into a linear scale ranging from 0 to 50. Here, 0 corresponded to 50% (indicating no preference, i.e., zero confidence), and 50 corresponded to 100% (reflecting maximal confidence in either decision). In cases where directionality was relevant,

the sign was used to indicate the decision's direction (i.e., which of the two options was chosen), effectively extending the linear scale to $[-50; 50]$.

The main effects of the background noise on the three decision-making measures were evaluated using chi-squared tests (two-tailed), and results are presented in contingency tables along with the 95% confidence intervals of the odds ratio of the two conditions. Interaction effects between noise and confidence ratings on the decision-making measures were estimated using logistic regression models, where pre-conversation confidence ratings were used as the predictor variable. Statistical significance of the difference between conditions was evaluated based on two-sided permutation test of the difference in the intercept (β_0) and slope (β_1) parameters of these models. The permutation test was performed by randomizing the loud and soft noise labels of each trial $N_p = 10000$ times, thus simulating N_p draws from the proposed null distribution where the condition has no effect on the outcome. Confidence intervals for the difference in parameter estimates between conditions are also reported; these were determined using the percentile bootstrap method using $N_b = 10000$ bootstrap samples for each condition. The bootstrapped samples from each condition were subtracted from each other, and the 5th and 95th percentiles of the N_b bootstraps are reported as the confidence intervals of the difference between conditions. Additionally, to account for potential confounds from individual biases at the group or subject levels, generalized linear mixed-effects models (GLMMs) were also fit to predict the decision-making measures. The details of these models are provided in the relevant context.

Results

Questionnaire Responses

To verify that the noise condition was challenging for the participants to communicate in, the difference in questionnaire responses was compared between the loud and soft noise conditions. On the 11-point scale, effort was rated much higher in the loud noise condition (mean difference, $\mu = 5.43$, 95th percentile bootstrapped CI: [5.01, 5.84], Cohen's $d = 3.58$), while ease of understanding ($\mu = -3.80$, CI:[$-4.28, -3.32$], $d = -2.20$), ease of expression ($\mu = -2.83$, CI: [$-3.35, -2.31$], $d = -1.65$), and conversation flow ($\mu = -1.87$, CI: [$-2.32, -1.41$], $d = -1.10$) were rated

lower in loud noise. Engagement was rated slightly lower in the loud noise condition ($\mu = -0.95$, CI: [$-1.44, -0.48$], $d = -0.48$). Taken together, these responses indicate that the experimental setup successfully created a challenging communication scenario in the loud noise condition.

Impact of Noise and Initial Confidence Rating on Stay/ Switch Behavior

Pairwise stay/switch behavior was first analyzed in relation to the loud and soft noise conditions and whether the deciding member was the most or least confident member. Trials were excluded if either member submitted an initial confidence rating of 0 (i.e., no preference). Additionally, when assessing the difference between the most and least confident members, trials were omitted if both participants submitted equal initial confidence ratings. Table 1 illustrates the stay/ switch behavior of the participants. Participants were more inclined to stay with their initial decision (65.6%) in trials where they were more confident than another member. Conversely, when they were the least confident member, they stayed with their initial response in only 42.8% of trials. There was no significant main effect of background noise on the likelihood of participants staying with their initial decision. A two-sided binomial test revealed a small but significant overall bias, with 54.2% of trials resulting in participants choosing to stay.

To explore the interaction between initial confidence ratings and background noise, each instance of pairwise disagreement was categorized based on stay/switch outcomes and loud/soft noise conditions. The left panel of Figure 2 illustrates these four trial categories with semitransparent markers, where each marker corresponds to a single pairwise prior disagreement. On the x-axis, the difference in the initial confidence rating between the two members in each trial is shown. Positive values indicate that the first-person member (the decision-maker) is more confident, while negative values suggest the opposite. The solid lines represent logistic regressions conducted separately in each condition, with β_1 denoting the slope and β_0 the intercept at $c_{me} - c_{you} = 0$. These regressions predict the probability of a trial being a *stay* trial as a function of the difference in initial confidence ratings. In both conditions, the slope was positive, consistent with the analysis from the most/least confident split in Table 1. This indicates that more confident

Table 1. Pairwise Stay/Switch Behavior by Noise Condition and by Relative Initial Confidence Rating.

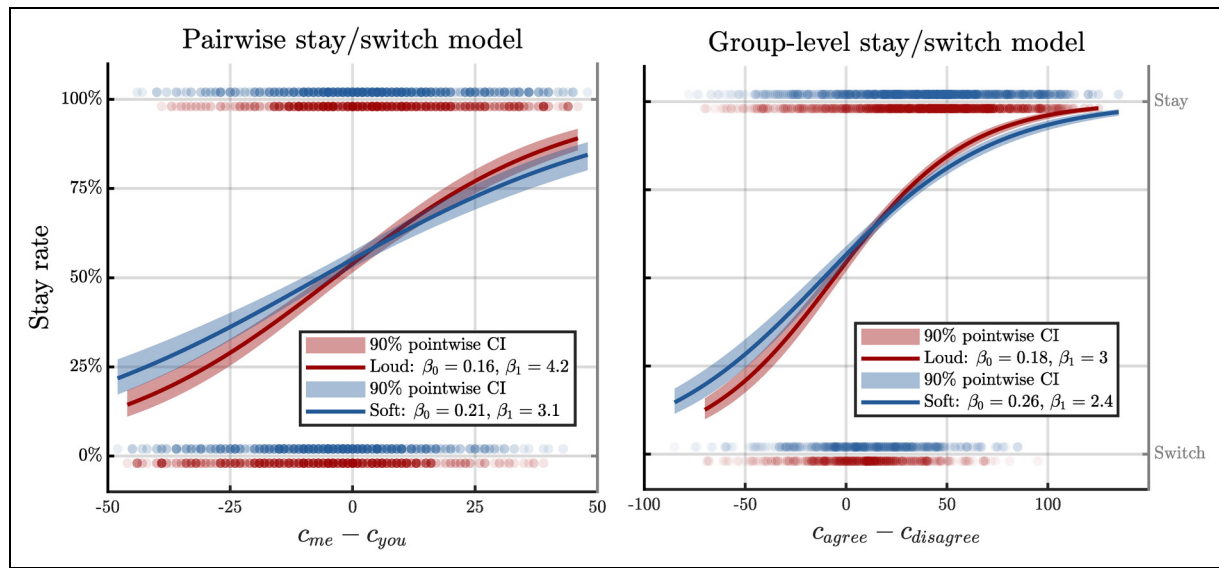|	# decisions	Stay	Switch	χ^2	p-value	Odds ratio [95% CI]
Most confident	1327	65.6%	34.4%	139.4	1e-32	2.55 [2.18;2.98]
Least confident	1327	42.8%	57.2%			
Loud (78 dB)	1284	53.5%	46.5%	0.433	0.511	0.95 [0.81;1.10]
Soft (48 dB)	1382	54.8%	45.2%			
All	2666	54.2%	45.8%	–	1.55e-5	–

Figure 2. Interaction between stay/switch behavior and the difference in initial confidence ratings of a pair or a group. Left panel: Stay/switch behavior as a function of the initial confidence rating difference between two disagreeing members. Positive values mean that the member making the decision to stay or switch provided the highest confidence rating. The slope of the regression is significantly steeper in the noise condition ($p = 0.0261$, $CI = [0.286; 1.957]$). There was no significant difference between the intercepts ($p = 0.548$, $CI = [-0.186; 0.0860]$). Right panel: Stay/switch behavior as a function of the sum of initial confidence ratings of those agreeing with the deciding member minus the sum of initial confidence ratings of those disagreeing. Again, the slope is significantly steeper in the noise condition ($p = 0.0044$, $CI = [0.261; 1.026]$), but there is no significant difference between the intercepts ($p = 0.363$, $CI = [-0.0247; 0.0675]$).

members are more likely to stay, whereas less confident members are more inclined to switch. The overall bias towards staying is evident in the y-axis intercept, which exceeds the 50% mark in both conditions.

The noise condition exhibited a significantly steeper slope in the regression models (pairwise model: $p = 0.0261$, $CI = [0.286; 1.957]$). This finding suggests an interaction between condition and initial confidence rating, indicating that in the noise condition, the prior answer of the most confident member was even more likely to be chosen after the conversation. Note that the slope parameters reported in the legend of Figure 2 refer to the log-odds domain regression. When translated to percentages, the slope at the origin is 1.05 for the noise condition and 0.77 for the soft noise condition. Thus, transitioning from equal confidence ratings ($c_{me} - c_{you} = 0$) to a one-point difference ($c_{me} - c_{you} = 1$) results in a 1.05 percentage point change in the stay/switch decision in noise, compared to a 0.77 percentage point change in soft noise.

The pairwise stay/switch model does not consider the contribution of the third group member, who may also influence the decision on any given trial. To address this, a model using the sum of initial confidence ratings was employed, shown in Figure 2 (right panel). Here, the x-axis shows the sum of signed confidence ratings on any given trial, with a positive sign for each member agreeing with the deciding member and negative sign for those disagreeing. For example, in a trial with prior decisions [A B B], and confidence ratings [18 33 21],

when predicting the first member's stay rate, the total confidence difference would be $c_{agree} - c_{disagree} = 18 - (33 + 21) = -36$. Conversely, when predicting the second and third members' stay rates, $c_{agree} - c_{disagree} = (33 + 21) - 18 = 36$. This model's prediction was applied to all trials, not just those with prior disagreement. The potential range of $c_{agree} - c_{disagree}$ was thus $[-99; 150]$, since the confidence rating of the first-person member is, by definition, always positive. The model yielded results similar to the pairwise difference model concerning the slope and intercept parameters. There was a slight bias towards staying at a total confidence difference of zero, and the loud noise condition slope was significantly steeper than the soft noise condition slope ($p = 0.0044$, $CI = [0.261; 1.026]$).

While the pairwise and group-level stay/switch models illustrated in Figure 2 demonstrate an interaction between relative confidence ratings and stay/switch decisions, they assume that the decisions made by participants are unit samples, which may be an inappropriate assumption given that decisions made by the same participant or within the same group may be correlated. Furthermore, by using the confidence difference as the predictor, the model assumes that confidence ratings of disagreeing members have an equal but opposite effect on stay/switch behavior. To address these limitations, a GLMM was fitted to the data of the group-level stay/switch model shown in the right panel of Figure 2. The stay/switch response data was modelled as a binomial distribution with a logit link function. Separate predictors were included for c_{agree} and $c_{disagree}$, as well as

random intercepts and slopes for each of the thirty subjects and for each of the ten groups. Random intercepts were also included for the interaction between group and condition, as well as between subject and condition. The condition variable was coded as 0 for the soft noise condition and 1 for the loud noise condition, This analysis revealed a significant positive main effect of c_{agree} on the decision to stay (main effect estimate: 10.2e-2, CI: [7.56e-2; 12.9e-2], $p = 7.53e-14$, $t(4488) = 7.50$)., but no significant main effect of $c_{disagree}$ on the decision to stay (main effect estimate: 1.27e-2, CI: [−4.69e-2; 2.15e-2], $p = 0.465$, $t(4488) = −0.730$). Likewise, a significant effect of the interaction with condition was found for c_{agree}, with the effect of c_{agree} on the decision to stay being larger in noise (interaction effect estimate: −2.77e-2, CI: [−4.66e-2; −0.880e-2], $p = 0.00409$, $t(4488) = −2.87$). No such interaction was found for $c_{disagree}$ (interaction effect estimate: 0.127e-2, CI: [−2.07e-2; 1.82e-2], $p = 0.898$, $t(4488) = −0.128$), This indicates that the interaction between initial confidence ratings and noise level is robust to the inclusion of individual biases, but that this interaction appears to be mainly driven by the confidence of agreement decisions rather than that of disagreement decisions.

Impact of Noise on Convergence

The number of convergent trials – i.e., trials where prior disagreement turned into agreement after the discussion – in each condition is presented in Table 2, both for pairwise and group-level convergence. Convergence was more prominent in the noise condition, with odds ratios of 1.61 and 1.90 for groups and pairs, respectively, suggesting that participants were more inclined to reach agreement on their post-conversation decisions in the noise condition. To account for potentially confounding effects of individual differences between groups and subjects, a GLMM was fit to the group-level data. Decision convergence was used as a binary outcome (coding convergent trials as 1) with a logit link and random effects for group and condition were included, as well as an interaction between these. This revealed a significant effect of the interaction between condition and convergence (interaction effect estimate: −0.417, CI: [-0.789; −0.0462], $p = 0.0276$, $t(693) = -2.21$), indicating higher likelihood of convergence in the loud noise condition.

The effect of initial confidence ratings on convergence at the population level is shown in Figure 3. Pairwise and group-level confidence differences were calculated in a similar way as in the stay/switch analysis. The sign of the resulting confidence difference was omitted, as it had served only to identify the decision direction of the deciding member. For the convergence measure, there is no single deciding member, as convergence is inherently a group-level measure. In the x-axis labels of Figure 3, $|c_1 − c_2|$ denotes the difference between the initial confidence ratings of the two members in question, and $|c_A − c_B|$ denotes the sum total of initial confidence ratings of those favoring option A minus that of those favoring option B. Both the pairwise and group-level models exhibit a steeper slope in loud noise than in soft noise, although the difference was not significant for either model (pairs: $p = 0.204$, $CI = [−0.0058; 0.0441]$, groups: $p = 0.250$, $CI = [−0.0054; 0.0443]$). Nevertheless, especially for the pairwise model, the 90% confidence bounds of the convergence rate are distinctly separated for all but the lowest confidence differences.

Impact of Noise on Voting Strategy

The analysis of voting strategies focused specifically on trials where the most confident participant was in the minority, constituting 9.4% of the total trials (157 out of 1680). As each of these trials contributed three individual posterior decisions, a total of 471 binary decisions were used for the voting strategy analysis. The findings, presented in Table 3, suggest a significant effect of the condition, with the noise condition leading to a higher use of the confidence slating strategy. In the soft noise condition, both strategies were equally probable, whereas in noise, 58.6% of decisions (130 out of 222) used the confidence slating strategy. To account for potentially confounding effects of individual differences between groups and subjects, a GLMM was fitted to the data. Voting strategy was used as a binary outcome (coding confidence slating as 0 and majority voting as 1) with a logit link and random effects for participant and condition was included, as well as an interaction between these. This revealed a significant effect of the interaction between condition and voting strategy (interaction effect estimate: 0.393, CI: [0.0137;0.773], $p = 0.0423$, $t(469) = 2.03$), indicating more confidence slating in the loud noise condition.

Figure 4 shows the choice of strategy as a function of the confidence difference. The confidence difference was computed as the sum of the initial confidence ratings of majority members ($c_{majority}$) minus the initial confidence rating of the most confident member ($c_{confident}$). The data indicate that the

Table 2. Convergence Rate in Loud and Soft Noise.

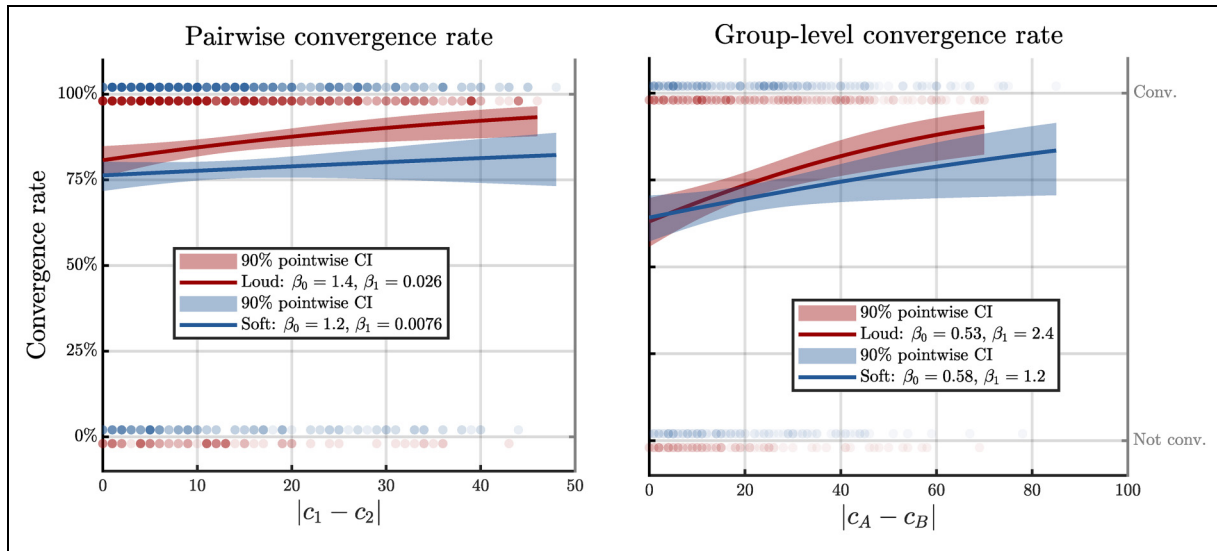|		# trials	Convergent	Non-convergent	χ^2	p-value	Odds ratio [95% CI]
Pairs	Loud	617	88.8%	11.2%	16.26	5.50e-5	1.90 [1.39; 2.61]
	Soft	668	80.7%	19.3%			
Groups	Loud	332	81.9%	18.1%	6.56	0.0108	1.61 [1.12; 2.31]
	Soft	363	73.8%	26.2%			

Figure 3. Interaction between initial confidence ratings and convergence rate. Left panel: Pairwise convergence as a function of the difference in initial confidence ratings for a disagreeing pair. Right panel: Group convergence as a function of the difference in the sum of initial confidence ratings between those favoring option A and those favoring option B. Although both slopes are steeper in noise, the difference is not statistically significant (pairs: $p = 0.204$, $CI = [-0.0058;\ 0.0441]$, groups: $p = 0.250$, $CI = [-0.0054;\ 0.0443]$). The intercepts are also not significantly different (pairs: $p = 0.245$, $CI = [-0.127;\ 0.641]$, groups: $p = 0.849$, $CI = [-0.119;\ 0.651]$).

Table 3. Voting Strategies in Loud and Soft Noise.

	# decisions	Majority voting	Confidence slating	χ^2	p-value	Odds ratio [95% CI]
Loud (78 dB)	222	41.4%	58.6%	3.96	0.0465	1.45 [1.01; 2.08]
Soft (45 dB)	249	50.6%	49.4%			

tendency for confidence slating in loud noise was most pronounced in trials where the most confident member's confidence was larger than the total confidence of the majority (negative values of $c_{majority} - c_{confident}$). There is substantial overlap between the regression confidence bounds, however, so this observation warrants further investigation. Moreover, due to sparse data at positive confidence differences, the observed effect of increased confidence slating in loud noise may be valid only for negative values of $c_{majority} - c_{confident}$. It remains uncertain whether majority voting might be more favored in noisy conditions when the total confidence of majority members is substantially larger than that of the minority member, i.e., when $c_{majority} - c_{confident} \gg 0$.

Discussion

The stay/switch behavior, convergence and voting strategy used in this study illustrate how interlocutors may influence each other in a collaborative, task-oriented conversation. The results from all three decision-making measures – stay/switch behavior, convergence and voting strategy – consistently indicated a similar effect: posterior decisions were more likely to me modified towards prior decisions made with high confidence, and this effect was more pronounced

in the loud noise condition. This was particularly evident in the voting strategy analysis, demonstrating a stronger inclination to align with high-confidence members in the noise condition, even when they were in the minority. In the stay/switch analysis, the interaction between noise level and confidence was revealed by the steeper slope in loud noise. The convergence rate also increased in loud noise, especially when one member showed substantially higher confidence than others or when the difference in total confidence for the two options was large.

It is not surprising that the analyses of confidence slating/majority voting and stay/switch decisions yielded similar results. In fact, the pairwise confidence difference model of stay/switch behavior can be viewed as a parametrized generalization of the binary CS/MV models. In the pairwise confidence difference model, an infinitely steep slope (i.e., a step function at $c_{me} - c_{you} = 0$) corresponds to the CS model of dyadic decision-making, where the decision is always determined by the most confident member. In an infinitely steep confidence difference model, for $c_{me} - c_{you} < 0$, the decision will be switch with 100% certainty, and for $c_{me} - c_{you} > 0$, the decision will be stay with 100% certainty. Conversely, if the slope is zero, stay and switch decisions are equally probable regardless of confidence levels – analogous to the
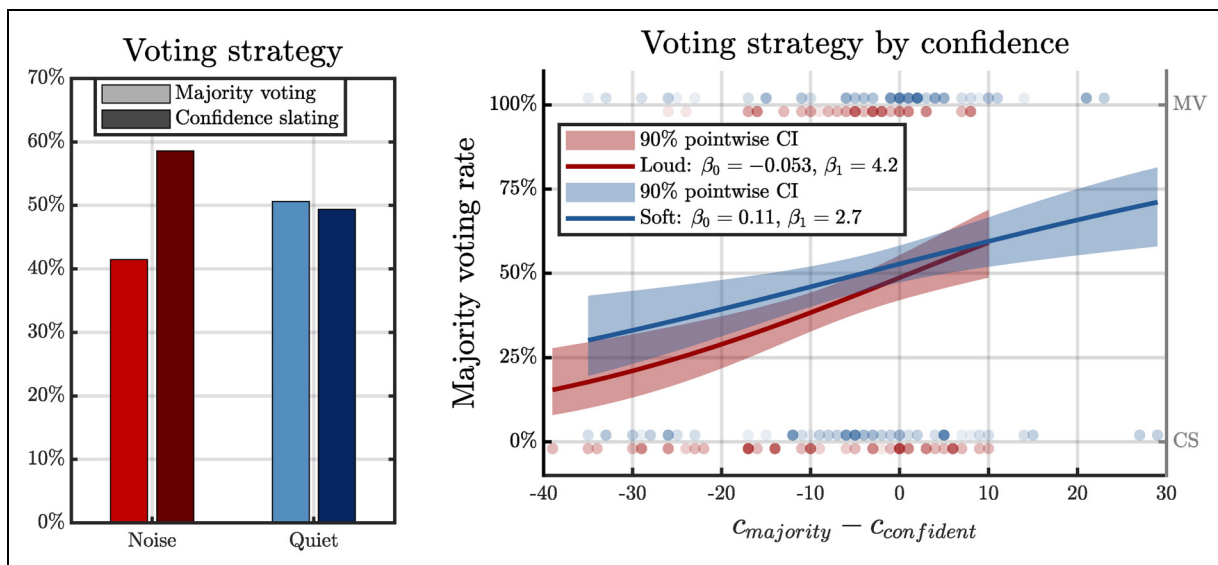
Figure 4. Left panel: overall prevalence of confidence slating and majority voting strategies. Right panel: Prevalence of CS and MV strategies as a function of the difference between the sum of initial confidence ratings of majority members and that of the most confident member. At negative values, the most confident member is more confident than the sum of the majority members' confidence. Only trials where the majority members and most confident member chose different options prior to the conversation are included.

MV model, which considers only the fact that there is one vote for each option and disregards confidence levels. Therefore, any positive finite slope represents an operating point between these two extremes, MV and CS, and a larger positive slope in loud noise can be interpreted as a stronger tendency towards CS.

Like the voting strategy, the convergence rate is also intricately connected to stay/switch behavior. To illustrate this, consider the example where two disagreeing group members have confidence levels of 36 and 11, respectively. The first member's confidence difference ($c_{me} - c_{you}$) is 25, while the second member's confidence difference is -25. Referring to Figure 2, at these confidence levels, member one would have a 77% chance of staying in the loud noise condition, as opposed to only 73% in the soft noise condition. Conversely, member two would have a 72% chance of switching in loud noise, compared to only 64% in soft. As pairwise convergence requires one member to stay and the other switch, the total probability of convergence would be $0.77 * 0.72 + (1 - 0.77) * (1 - 0.72) = 0.62$ in loud noise, but only $0.73 * 0.64 + (1 - 0.73) * (1 - 0.64) = 0.56$ in soft. In fact, at any given confidence difference level, if the stay rate at positive confidence differences is higher in loud noise and the stay rate at negative confidence differences is lower, the expected convergence rate will similarly be higher in loud noise, as both members are more likely to choose the same outcome.

What could be the underlying reason that confident members seem to have more influence on posterior decisions in the loud noise condition? We speculate that the underlying reason for this might be that not all disagreements are actually

verbalized in the conversation, and that the decision to verbalize a disagreement depends on both the noise level and the initial confidence ratings of individuals. To explain how this could promote decisions that rely mostly on members who are more confident, consider a trial where the group submitted disagreeing decisions prior to the conversation. During the conversation, each group member must individually decide whether to try to convince others that their belief is correct. While ideally, all opinions would be shared and weighted based on the degree of confidence with which they are held – this has been shown in Grofman et al. (1983) to lead to ideal decision-making under certain assumptions – this is not always the case. In loud background noise, low-confidence opinions may be at particular risk of not being voiced, as their authors may find their own conviction too low when weighed against the effort required to express it. Expressing an opinion may lengthen the conversation unnecessarily, especially if an opposing opinion has already been expressed. If one's opinion is not held with high confidence, the potential reward for expressing it (i.e., increased task performance for the group) is low. The decision to share an opinion could thus be influenced by an interaction between the confidence level and the background noise level. It seems plausible that opinions held with greater confidence are more likely to be shared, regardless of noise level, as they could be perceived to have a greater potential for increasing task performance. If the decision on whether to voice low-confidence opinions is more sensitive to the increased effort requirement, high-confidence opinions would become relatively more likely to go unchallenged when the noise level increases. This would explain the stronger influence that high-confidence individuals

seem to have on posterior decisions in noise. While the present study contains no direct evidence for this explanation, there is some indirect support for it in the duration of the conversations recorded. Groups discussed for on average 8.0 min in the soft noise condition but only 6.8 min in the loud noise condition, a significant difference ($p = 0.0145$, $CI = [-1.836; -0.4206]$). This observation aligns with the hypothesis that fewer disagreements were voiced in loud noise, although it should be noted that other effects, such as increased speech rate in noise, may also account for this difference.

The proposed explanation for why confidence matters more in loud noise relies on two core assumptions: 1) that communicating in loud background noise requires more effort, and 2) that participants' initial confidence ratings are related to how important they find it to voice their opinion in order to increase the group's task performance. The first assumption is a fairly well-established phenomenon; noise levels similar to or lower than those used in the present study have previously been shown to be rated as highly effortful to communicate in (Beechey et al., 2018; Hazan et al., 2019). The second assumption is less clearly fulfilled in the present study. While participants were presented with a score from their initial decisions at the end of each round, this score was not tied to the confidence rating. There was thus no reward-based incentive promoting accurate confidence ratings. Moreover, postponing the reward until the end of the round might reduce its efficiency in terms of promoting confidence ratings that relate to perceived importance of one's decision. Providing a more immediate reward structure tied to the initial confidence ratings might be necessary to fully justify this assumption.

In the present study, any direct observation of whether the suggested tradeoff between effort and importance discourages individuals from expressing disagreements on specific questions was hindered by the arrangement of questions into lists of transitively related items. Due to the free nature of the discussions, and because groups often resorted to ranking the items instead of discussing them pair-by-pair, effectively labeling each spoken statement as being connected to a particular item pair would prove challenging, if not impossible. To rigorously test the hypothesized explanation proposed here, future studies might benefit from avoiding the list-based experimental setup and instead isolate each decision into a separate discussion. This way, it could be directly observed whether low-confidence opinions are less likely to be voiced in loud levels of background noise. This would also increase the internal validity of the findings, as initiating the discussions immediately after the confidence rating would ensure better coherence between confidence ratings and level of expressed confidence in the conversation. Investigating whether submitted confidence ratings and expressed confidence are related would also be facilitated by an experimental setup that is based on discussions about single item pairs rather than lists.

In a similar vein, it might also be necessary to further investigate the assumption that initial confidence ratings reflect the expressed confidence of individuals. Contrary to expectations,

the statistical analysis revealed a main effect of c_{agree} on stay/switch decisions, but not of $c_{disagree}$. Participants who submitted lower initial confidence ratings were more likely to switch and vice versa, but the confidence of other members did not significantly impact the decision. This suggests that participants were influenced by their own prior metacognitive judgments and incorporate these into their posterior decisions, but that other members' expressed confidence was not interpreted on a continuous scale. This would contrast with previous findings, which have suggested that people are able to accurately make use of each other's metacognitive judgments to improve decision making, although these findings related specifically to dyads performing a joint perceptual task (Bahrami et al., 2010).

Regardless of whether discussions are based on lists or individual item pairs, using pre- and post-conversation responses to investigate if and how group members influence each other's decisions has its limitations. While it may be tempting to conclude that any change in individual decisions is due to influence from other group members, variation in responses within participants can occur even without interaction with others. Confidence ratings are susceptible to internal noise, as metacognitive judgements can be difficult to make accurately (Bang et al., 2014; Gigerenzer et al., 1991; Kleitman & Stankov, 2001). The convergence measure is especially sensitive to this, as low confidence-opinions may more easily shift from one decision to the other, potentially resulting in false positive convergence trials due to chance. It should be emphasized, however, that the convergence rates observed in this study are much higher than what would be expected by chance alone, which in the most extreme case one low-confidence member disagreeing with two high-confidence members would be 50%. There may also be a bias related to seeing the same question multiple times, such that, for example, confidence ratings might increase simply because the question has been answered recently (i.e., before the conversation). These sources of error might be alleviated by including control trials that are not related to the items discussed, but which nonetheless require participants to make metacognitive judgments before and after the conversation.

The framework introduced here offers a high degree of flexibility in implementation. Firstly, the stay/switch, convergence and voting strategy measures can generally be derived regardless of group size, especially the group-level variants of the measures. This represents a clear advantage over, for example, the commonly used diapix task, which is inherently limited by its task material to dyads. Additionally, the general knowledge questions can naturally be replaced by questions from arbitrary knowledge domains. They can also be substituted with entirely different types of questions. For example, a 2-alternative perceptual task or even opinion-based questions with no correct answer could be employed. Both of these alternative question types have been used in the group decision-making literature (Bang et al., 2014; Fay et al., 2000; Koriat, 2012), and the analysis tools introduced

in this study can be directly applied to decisions using those kinds of questions as well. Different question types will undoubtedly yield different decision-making behaviors, and the impact of noise, hearing impairment, or other communicative barriers should be considered in the context of which type of question is being used.

Conclusion

This study introduced a novel task designed for group interaction studies. This task, applicable to any group size, used a decision-making paradigm with general-knowledge questions to shed light on how individual group members modified their initial decisions after conversing in two different levels of background noise. Experiments with ten normal-hearing triads revealed significant differences in decision-making behavior based on the background noise level during conversations. After conversing in loud background noise, posterior decisions were more likely to change towards prior decisions held with high confidence. In contrast, when the conversations took place in soft noise, prior decision confidence seemed to play a smaller role in determining the posterior decisions. The task paradigm used in this study is widely used in the scientific literature on group decision-making, and we believe that it holds considerable promise in hearing research, where it could provide insights into how hearing impairment affects individuals' participation in – and influence on – group decision-making processes.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iDs

Ingvi Örnolfsson https://orcid.org/0000-0002-2222-0739
Torsten Dau https://orcid.org/0000-0001-8110-4343
Tobias May https://orcid.org/0000-0002-5463-5509

Data Availability Statement

The confidence and decision data gathered from participants in this study is publicly available from DTU Data at DOI 10.11583/DTU.25163816 (Örnolfsson et al., 2024).

References

Baker, R., & Hazan, V. (2011). DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, *43*(3), 761–770. https://doi.org/10.3758/s13428-011-0075-y

Bang, D., Fusaroli, R., Tylén, K., Olsen, K., Latham, P. E., Lau, J. Y. F., Roepstorff, A., Geriant, R, Frith, C. D., & Bahrami, B. (2014). Does interaction matter? Testing whether a confidence heuristic can replace interaction in collective decision-making. *Consciousness and Cognition*, *26*(100), 13–23. https://doi.org/10.1016/j.concog.2014.02.002

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, *79*(6), 941–952. https://doi.org/10.1037/0022-3514.79.6.941

Beechey, T., Buchholz, J. M., & Keidser, G. (2018). Measuring communication difficulty through effortful speech production during conversation. *Speech Communication*, *100*, 18–29. https://doi.org/10.1016/j.specom.2018.04.007

Beechey, T., Buchholz, J. M., & Keidser, G. (2019). Eliciting naturalistic conversations: A method for assessing communication ability, subjective experience, and the impacts of noise and hearing impairment. *Journal of Speech, Language, and Hearing Research*, *62*(2), 470–484. https://doi.org/10.1044/2018_JSLHR-H-18-0107

Berlo, D. K. (1960). *The process of communication: an introduction to theory and practice*. Holt, Rinehart and Winston, Inc.

Buchholz, J. M., Davis, C., Beadle, J., & Kim, J. (2022). Developing a real-time test to investigate conversational speech understanding. *Journal of Speech, Language, and Hearing Research*, *65*(12), 4520–4538. https://doi.org/10.1044/2022_JSLHR-22-00218

Carlile, S., & Keidser, G. (2020). Conversational interaction is the brain in action: Implications for the evaluation of hearing and hearing interventions. *Ear & Hearing*, *41*(Supplement 1), 56S–67S. https://doi.org/10.1097/AUD.0000000000000939

Fay, N., Garrod, S., & Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological Science*, *11*(6), 481–486. https://doi.org/10.1111/1467-9280.00292

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science*, *23*(8), 931–939. https://doi.org/10.1177/0956797612436816

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, *1*(2), 292–304. https://doi.org/10.1111/j.1756-8765.2009.01020.x

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*(4), 506–528. https://doi.org/10.1037/0033-295X.98.4.506

Grofman, B., Owen, G., & Feld, S. L. (1983). Thirteen theorems in search of the truth. *Theory and Decision*, *15*(3), 261–278. https://doi.org/10.1007/BF00125672

Hadley, L. V., Brimijoin, W. O., & Whitmer, W. M. (2019). Speech, movement, and gaze behaviours during dyadic conversation in noise. *Scientific Reports*, *9*(1), 10451. https://doi.org/10.1038/s41598-019-46416-0

Hadley, L. V., & Ward, J. A. (2021). Synchrony as a measure of conversation difficulty: Movement coherence increases with background noise level and complexity in dyads and triads.

PLOS ONE, *16*(10), e0258247. https://doi.org/10.1371/journal.pone.0258247

Hadley, L. V., Whitmer, W. M., Brimijoin, W. O., & Naylor, G. (2021). Conversation in small groups: Speaking and listening strategies depend on the complexities of the environment and group. *Psychonomic Bulletin & Review*, *28*(2), 632–640. https://doi.org/10.3758/s13423-020-01821-9

Hazan, V., Tuomainen, O., & Taschenberger, L. (2019). Subjective evaluation of communicative effort for younger and older adults in interactive tasks with energetic and informational masking. In *Interspeech 2019* (pp. 3098–3102). ISCA. Retrieved 28 March 2023 from: https://doi.org/10.21437/Interspeech.2019-2215

Keshmirian, A., Deroy, O., & Bahrami, B. (2022). Many heads are more utilitarian than one. *Cognition*, *220*, https://doi.org/10.1016/j.cognition.2021.104965

Kiessling, J., Pichora-Fuller, M., Gatehouse, S., Stephens, D., Arlinger, S., Chisolm, T., & Wedel, H. (2003). Candidature for and delivery of audiological services: Special needs of older people. *International Journal of Audiology*, *42*(Suppl 2), 2S92–2101. https://doi.org/10.3109/14992020309074650

Kleitman, S., & Stankov, L. (2001). Ecological and person-oriented aspects of metacognitive processes in test-taking. *Applied Cognitive Psychology*, *15*(3), 321–341. https://doi.org/10.1002/acp.705

Koriat, A. (2012). When are two heads better than one and why? *Science*, *336*(6079), 360–362. https://doi.org/10.1126/science.1216549

Lu, H., McKinney, M. F., Zhang, T., & Oxenham, A. J. (2021). Investigating age, hearing loss, and background noise effects on speaker-targeted head and eye movements in three-way conversations. *The Journal of the Acoustical Society of America*, *149*(3), 1889–1900. https://doi.org/10.1121/10.0003707

Mahmoodi, A., Bahrami, B., & Mehring, C. (2018). Reciprocity of social influence. *Nature Communications*, *9*(1), 2474. https://doi.org/10.1038/s41467-018-04925-y

Meyen, S., Sigg, D. M. B., Luxburg, U. v., & Franz, V. H. (2021). Group decisions based on confidence weighted majority voting. *Cognitive Research: Principles and Implications*, *6*(1), 18. https://doi.org/10.1186/s41235-021-00279-0

Miles, K., Weisser, A., Kallen, R. W., Varlet, M., Richardson, M. J., & Buchholz, J. M. (2023). Behavioral dynamics of conversation, (mis)communication and coordination in noisy environments. *Scientific Reports*, *13*(1), 20271. https://doi.org/10.1038/s41598-023-47396-y

Nicoras, R., Gotowiec, S., Hadley, L. V., Smeds, K., & Naylor, G. (2022). Conversation success in one-to-one and group conversation: A group concept mapping study of adults with normal and impaired hearing. *International Journal of Audiology*, *62*(9), 868–876. https://doi.org/10.1080/14992027.2022.2095538

O'Connell, D. C., Kowal, S., & Kaltenbacher, E. (1990). Turn-taking: A critical analysis of the research tradition. *Journal of Psycholinguistic Research*, *19*(6), 345–373. http://dx.doi.org/10.1007/BF01068884

Örnolfsson, I., Ahrens, A., Dau, T., & May, T. (2024). *The effect of collaborative triadic conversations in noise on decision-making in a general-knowledge task*. Technical University of Denmark. Dataset. https://doi.org/10.11583/DTU.25163816.v1

Peperkoorn, L. S., Becker, D. V., Balliet, D., Columbus, S., Molho, C., & Van Lange, P. A. M. (2020). The prevalence of dyads in social life. *PLoS ONE*, *15*(12), e0244188. https://doi.org/10.1371/journal.pone.0244188

Petersen, E. B. (2024). Investigating conversational dynamics in triads: Effects of noise, hearing impairment, and hearing aids. *Frontiers in Psychology*, *15*, 761. http://dx.doi.org/10.3389/fpsyg.2024.1289637

Petersen, E. B., MacDonald, E. N., & Josefine Munch Sørensen, A. (2022). The effects of hearing-aid amplification and noise on conversational dynamics between normal-hearing and hearing-impaired talkers. *Trends in Hearing*, *26*. http://dx.doi.org/10.1177/23312165221103340

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232. http://dx.doi.org/10.1016/0010-0285(89)90008-X

Sørensen, A. J. M., Fereczkowski, M., & MacDonald, E. N. (2021). Effects of noise and second language on conversational dynamics in task dialogue. *Trends in Hearing*, *25*. http://dx.doi.org/10.1177/23312165211024482

Watson, S., Sørensen, A. J. M., & MacDonals, E. N. (2019). The effect of conversational task on turn taking in dialogue. Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR). 7. https://proceedings.i-saar.eu/index.php/isaarproc/article/view/2019-08