How awareness of each other's mental load affects dialogue

Dominique Knutsen, Angele Brunelliere

How awareness of each other's mental load affects dialogue

Dominique Knutsen and Angèle Brunellière

Univ. Lille, CNRS, CHU Lille, UMR 9193 - SCALab - Sciences Cognitives et Sciences Affectives, F-59000 Lille, France

Corresponding author:

Dr Dominique Knutsen

Laboratoire SCALab UMR CNRS 9193

Université de Lille

Rue du barreau

BP 60149, 59653 Villeneuve d'Ascq Cedex

France

Email address: dominique.knutsen@univ-lille.fr

The data and the analysis code can be downloaded at https://osf.io/6uqpv/

Acknowledgments:

We would like to thank Emma Augereau, Prescillia Bioche, Maréva Brunet, Sophie Carré, Guillaume Clabaut, Alicia Fasquel, Etienne Géhenne, Joséphine Lacroix, Louise Lefevre, Maëlle Legrand, Andréa Lorthioir and Amandine Thévenart for their help with the data collection and analysis. This work was supported by a MESHS "Structurants" grant (DEFLANG project) and the French National Research Agency (ANR-19-CE28-0006).

Abstract

During dialogue, people reach mutual comprehension through the production of feedback markers such as “yeah” or “okay”. The purpose of the current study was to determine if mental load affects feedback production, as there is currently no consensus as to how mental load constrains the way in which dialogue partners reach mutual comprehension. In two experiments, pairs of participants interacted in order to complete a collaborative puzzle game. We manipulated the amount of mental load experienced by each participant by giving them a series of digits to memorise (or no digits) before the beginning of the game. In Experiment 1, the participants were given no information about their partner’s mental load. In Experiment 2, each participant was told whether their partner had received digits to memorise. We found that although some results were identical in both experiments (directors produced more words, longer utterances and fewer feedback markers than matchers), the effect of mental load was different in both experiments. Indeed, whereas in Experiment 1, mental load mainly affected the number of words and speech turns produced, in Experiment 2, participants who had to follow the instructions of their partner and were under low mental load produced more feedback markers when their partner was under high mental load. Taken together, these results help disentangle the contribution of experienced and perceived mental load on collaboration in dialogue. They also highlight the importance of being explicitly aware of each other’s mental load in inter-personal coordination.

Keywords

Dialogue; mental load; feedback production; least collaborative effort

How awareness of each other's mental load affects dialogue

Despite its apparent ease, dialogue is a complex activity which requires two people (or more) to coordinate in order to reach mutual comprehension together (e.g., Brennan et al., 2010; H. H. Clark, 1996; Fusaroli et al., 2012; Garrod & Pickering, 2004). For instance, Clark and colleagues have suggested that people may use various kinds of signals to indicate that the addressee has understood what the speaker said well enough for current purposes (H. H. Clark & Brennan, 1991; H. H. Clark & Schaefer, 1989; H. H. Clark & Wilkes-Gibbs, 1986; see also Gergle et al., 2013; Lysander & Horton, 2012; McInnes & Attwater, 2004; Roque & Traum, 2008), a process referred to as grounding (Clark & Brennan, 1991). One available device to do so is "backchannels" feedbacks, that is, words or short expressions such as "yeah", "okay" or "I see". The expression "backchannel" is used to highlight the fact that the speaker producing such feedback markers does not intend to take the floor; rather, they intend to signal their understanding and active participation in the conversation (e.g., Bangerter & Clark, 2003; Knudsen et al., 2020; Tolins & Fox Tree, 2014).

Importantly, backchannel feedback markers may be used not only to ground information, but also to *navigate* the joint activities that both dialogue partners engage into as they interact. Past research highlights that feedback markers play an important role in dialogue success, insofar as they enable dialogue partners to make sure that they understand each other well enough and to clarify which topic or project is currently under discussion. It is important to highlight that although grounding and dialogue navigation are presented here (and in the literature) as two separate processes, in fine, they both serve the same purpose, that is, to enable both partners to reach mutual comprehension. For instance, Bangerter and H. H. Clark (2003) examined the use of feedbacks markers such as "yeah" and "okay" in situations where two people used dialogue to coordinate joint activities such as moving a bench or solving a puzzle together. They found that feedback markers tended to be used either to proceed within the current joint project (these markers, which include words such as “yeah” or “uh-huh”, are referred to as horizontal markers) or to enter and exit from projects (these markers, which include words such as “okay”, are referred to as vertical markers) (see also Bangerter et al., 2004, 2020; Knutsen et al., 2019). So, for instance, a speaker may say “uh-huh” to signal their continued attention while their partner explains how to place a specific piece of the puzzle; they may then say “okay” once the puzzle piece has been placed correctly, thus signalling their willingness to move on to the next joint project (e.g., figuring out where to place another piece of the puzzle).

Even though feedback markers are considered as playing an important role in mutual comprehension, little is known as to how the features of the situation in which the dialogue takes place may affect feedback production. This is surprising, as attempting to answer this question would help shed light on the determinants of inter-personal coordination in dialogue. In this context, the current paper specifically focuses on the impact of one feature of the dialogue context, namely the amount of mental load experienced by each participant as they interact. Mental load may be defined as the demand imposed by any task on a person’s limited mental resources (see Wickens, 2008; see also Hart & Staveland, 1988). Boiteau et al. (2014) have suggested that dialogue involves a fairly high mental load for all participants, in particular due to the demands of speaking. Indeed, in their study, the authors asked participants to interact with another person while tracking a target on a screen: they found that the participants’ performance on the tracking task was worse when the participants were talking than when they were listening to the other person.

Research has also shown that this mental load has a strong impact on how people communicate. For instance, mental load influences speech production, as it causes heart rate and vocal intensity to increase (e.g., Brenner et al., 1994). Mental load also appears to attenuate the use of acoustic details in speech perception, as listeners under high mental load rely more on lexical and semantic information when segmenting speech (Mattys et al., 2009, 2014; Mattys & Wiget, 2011). Interestingly, mental load also affects the speakers' ability to reach mutual comprehension. Indeed, Rossnagel (2000) found that people who experience high levels of mental load fail to take their partners' perspective into account in a dialogue. Indeed, in his study, participants were asked to provide instructions to a partner so that they could assemble a machine model. The partner was either an adult or an 8-year-old child. When also asked to memorise complex information and/or series of digits before interacting with their partner, the participants failed to tailor the content of their utterances depending on their partner's identity (e.g., they failed to avoid using technical terms when addressing an 8-year-old child rather than an adult). Additionally, Rossnagel (2004) found that people find it harder to take their partner's feedback into account when they experience high levels of mental load. In line with these findings, Abel and Babel (2017) showed that speech convergence (i.e., increasing similarity in speech production between dialogue partners) is less likely to occur when people experience high levels of mental load. Taken together, these studies suggest that mental load makes it more difficult for dialogue partners to reach mutual comprehension, as it makes people less likely to rely on perspective-taking and feedbacks. But in contrast with this idea, Knutsen et al. (2018) recently conducted a study in which a participant (the director) provided instructions to help another participant (the matcher) assemble the pieces of a tangram puzzle. They found that mental load (as well as face visibility) actually causes people to produce *more* feedback when under high levels of mental load. Based on this finding, the authors suggested that people attempt to "help" their partner under high mental load by producing more feedback, thus guiding their partner more explicitly through the interaction. This would enable both partners to reduce their collaborative effort, i.e. the total amount of effort put into interacting – although this assumption was not tested directly in the study. In sum, although *experiencing* mental load has a negative impact on some aspects of dialogue (i.e., perspective-taking and feedback perception), *perceiving* that one's partner is experiencing difficulties due to a high level of mental load may trigger helpful behaviours such as producing more feedback markers. In other words, whether mental load impairs or improves collaboration in dialogue depends on whether mental load is *experienced* or *perceived.*

However, there might be an alternative explanation to Knutsen et al.'s (2018) finding: the participants could have produced more feedback not because their *partner* experienced high levels of mental load, but because they *themselves* experienced high levels of mental load. Indeed, the methodology used in Knutsen et al.'s (2018) study implied that both partners experienced the same amount of mental load (high or low) at the same time, making it impossible to confirm that the positive influence of mental load on feedback production is triggered by the perception of one's partner's high mental load. The results from Rossnagel's (2000, 2004) studies cannot be used to answer this question either, as only one person experienced high levels of mental load (the main speaker in the 2000 study and the listener, who needed to interpret their partner's feedback, in the 2004 study). Disentangling the role of the mental load experienced by the speaker, on one hand, and the role of the partner's perceived mental load, on the other hand, is particularly important, as both partners might experience the same amount of mental load as they interact, but not always.

The question of whether feedback production depends on experienced or perceived mental load (or both) also has important theoretical implications. Indeed, feedback production depending mainly on one's own mental load would be in line with an egocentric approach to dialogue whereby early steps of language production are guided by one's own state of mind (e.g., Barr & Keysar, 2002; Epley et al., 2004). In contrast, feedback production depending mainly on the other person's mental load would be more in line with a collaborative approach to dialogue whereby language production is guided by speakers' efforts and attempts to reach mutual comprehension (e.g., Clark, 1996)

In this context, the purpose of the current study is to offer a better understanding of how mental load affects dialogue, by directly examining not only how one's own mental load affects feedback production, but also how the amount of mental load experienced by one's partner affects feedback production as well. Two experiments were conducted in which pairs of participants (i.e., dyads) jointly performed a collaborative puzzle-solving task. One of the participants (hereafter the director) was given the solution to the puzzle and had to provide instructions so that the other participant (hereafter the matcher) could solve it. Following Rossnagel (2000), the mental load of each participant was manipulated by asking them to remember a series of 7 digits (see also Baddeley & Hitch, 1974; Della Salla et al., 2010). Specifically, mental load was manipulated so that both participants experienced a high level of mental load, only one participant experienced a high level of mental load, or no participant experienced a high level of mental load.

In both experiments, our main hypothesis was that the effect of mental load on feedback production is driven by one's partner's *perceived* level of mental load, in line with Knutsen et al.'s (2018) suggestion and with the broader idea that experienced mental load affects people's ability to collaborate during dialogue. If verified, this would imply that participants should produce more feedback markers when their partner experiences high levels of mental load than when their partner does not experience high levels of mental load.

However, as highlighted above, previous research has shown that *experiencing* high levels of mental load prevents speakers from taking their partners' dialogic needs into account (i.e., they engage less in perspective-taking and fail to take their partners' feedback into account; Rossnagel, 2000, 2004). In line with this idea, speakers who experience high levels of mental load could fail to notice the amount of mental load experienced by their partners, thus preventing them from producing more feedback markers when needed. Thus, our second hypothesis is that noticing that one's partner is experiencing high levels of mental load may only be possible when one has sufficient mental resources to do so. In other words, participants should only produce more feedback markers when their partner experiences high levels of mental load when they are not under high mental load themselves.

We also sought to determine whether it is necessary for both partners to be explicitly aware of each other's levels of mental load to adapt feedback marker production. We raise this question because the hypotheses listed above are based on the assumption that people must be aware of their own mental load and their partner's mental load for the effect of mental load on feedback production to occur. However, because this question has seldom been addressed in the literature, it is unclear what "being aware of each other's mental load" actually means. One first possibility is that both partners must be given this information explicitly, as in Knutsen et al.'s (2018) study, in order to use it to adapt feedback production to their partner's perceived needs. However, information about one's dialogue partner's mental load is seldom explicitly provided in everyday conversation. Thus, a second possibility is that dialogue partners are able to infer each other's mental load based on a number of "generic" dialogic cues (e.g., speech rate, hesitations, fillers, etc.) and that explicitly stating the amount of mental load experienced by each partner is not necessary. In order to compare these two possibilities, participants in Experiment 1 were given no explicit information regarding their partner's mental load. We then asked them to try to assess their partner's mental load at the end of the experiment. As for participants in Experiment 2, they were explicitly told whether one of the partners (or both) experienced a high level of mental load. If people use generic dialogic cues to assess each other's mental load, we expect participants in Experiment 1 to provide an accurate estimate of their partner's mental load; we also expect to find a similar pattern of results in both experiments (i.e., feedback production should depend not only on experienced mental load, but also on perceived mental load, in both experiments). In contrast, if people need to be given explicit information about each other's mental load in order to be able to use this information, we would expect the effect of perceived mental load to be visible in Experiment 2 only. Please note that these hypotheses are necessarily exploratory because there is not enough literature on this topic (to our knowledge) and that comparing these hypotheses will provide a better understanding of the contribution of mental load to collaboration in dialogue.

Finally, in addition to examining how experienced and perceived mental load affect feedback production during dialogue, we also examined a number of additional variables in our study in order to determine how mental load also affects collaboration in dialogue. In order to do this, we chose variables which are often used in the literature to quantify the amount of effort put into the dialogue by both partners. In particular, we examined the number of words and speech turns produced by the participants, but also the time taken to complete the task and average utterance length (e.g., Bangerter et al., 2020; Clark & Wilkes-Gibbs, 1986).

Experiment 1

Participants

A total of 58 participants divided into 29 dyads took part in the experiment. Forty-three of them identified as female, 14 as male and one as non-binary. Fifteen dyads included only female participants, 12 dyads included one female and one male participant, one dyad included only male participants and one dyad included one female participant and one non-binary participant. The participants' average age was 20.34 years (*SD* = 3.59). All participants were native French speakers with no language disorders. They took part in the study for a small payment (10€). Within each dyad, the participants did not know each other prior to the beginning of the experiment.

All participants signed an informed consent form before the beginning of the experiment and were fully debriefed after the end of the study. This study was approved by the Ethics Committee of the University of Lille.

Apparatus

The dialogues between the participants were recorded using a double-entry voice recorder (Tascam DR-40).

Materials

The eight tangram figures which were used in this study were all made of the same seven pieces: one square, two small triangles, one medium triangle, two big triangles and one parallelogram. An additional tangram figure was used during the practice trial, but this figure was never used in the experimental trials. Two different versions of each figure were printed on separate A6 sheets. In the *director version*, the smaller pieces which made up the figure were apparent (see Figure 1a). In the *matcher version*, the whole figure was apparent but the smaller pieces which made it up were not (see Figure 1b). The seven loose pieces were also printed (see Figure 1c).

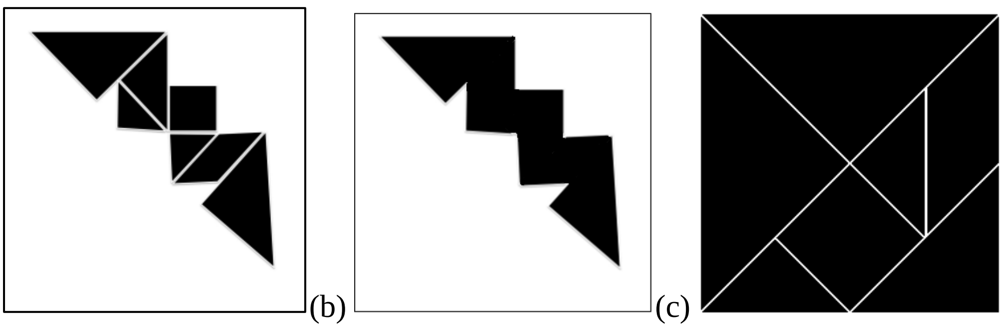
(a)


Figure 1. Example of one of the tangram figures used in the study. (a) Director's figure (indicating how to place the seven pieces in order to complete the puzzle). (b) Matcher's figure. (c) The seven pieces used by the matcher to assemble the puzzle.

At the beginning of each trial, each participant was given a small piece of paper which included either a series of digits to memorise (see Figure 2a) or a note stating that they had no digits to memorise during this trial (see Figure 2b). Ten series of seven digits were generated randomly.

(a)

Chiffres à mémoriser :

4 – 6 – 2 – 5 – 8 – 1 – 3

(b)

Vous n'avez pas de chiffres à mémoriser.

Figure 2. (a) Example of a series of digits to memorise. "Chiffres à mémoriser" means "Digits to memorise". (b) Note stating that the participant had no digits to memorise during the current trial. "Vous n'avez pas de chiffres à mémoriser" means "You have no digits to memorise".

After the end of each trial, the participants were asked to individually write down the digits they had memorised (or a series of seven zeros if they had not been given any digits to memorise) on an A5 piece of paper.

Task and procedure

The experiment took part in a quiet room of the psychology department of the University of Lille. The participants were seated at two different tables. Their chairs were positioned so that they sat back-to-back throughout the entire experiment, at a distance of approximately 30 centimetres. Thus, they could hear each other and interact naturally, but they could not see each other, preventing them from using non-linguistic cues (gestures, head nods, facial expressions, etc.) to communicate.

The participants started by signing an informed consent form. They were then told that they were about to perform a communication task together. They both played a different role in this task. The “director” was given the solution to a tangram figure, that is, they were given the version of the figure in which the seven pieces which made it up were apparent (see Figure 1a). As for the “matcher”, they were given the seven loose pieces which made up the figure (see Figure 1c) as well as a picture of the figure in which the seven pieces were not apparent (see Figure 1b).

More specifically, the director’s task was to give the matcher instructions to enable them to complete the figure, using the seven pieces they had been provided. To do so, the participants could interact freely: the director could talk as much as they wanted and the matcher could ask as many questions as they liked. The participants had up to five minutes to complete each figure. If they had finished before then, they could tell the experimenter and move on to the next trial. The experimenter interrupted them if they went over the five-minute limit. The participants repeated the task five times (one practice trial and four experimental trials). Their roles (director or matcher) remained the same throughout the entire experiment. After the end of each trial, the experimenter told the participants whether the matcher had completed the figure correctly. They were provided no additional information if they failed to complete the task (e.g., they were not told which pieces of the figure had been misplaced).

Before the beginning of each trial, the participants were each given a small piece of paper (see Figures 4 and 5) with information about the digits to memorise during the trial. They were instructed not to discuss the information shown on each sheet of paper, nor were they allowed to repeat the digits out loud during the trial. After the end of each trial, each participant individually wrote down the digits they remembered or a series of zeros (the reason why the participants were asked to write something down even in trials where they had no digits to memorise was to prevent their partner from guessing whether or not they had received any digits to memorise).

The figure used in the practise trial was always the same and neither participant had digits to memorise. The four pictures used in the four experimental trials were selected randomly and presented in a random order. During the experimental phase, the director was given digits to memorise in two trials and no digits to memorise in two trials; the matcher was also given digits to memorise in two trials and no digits to memorise in two trials. As mentioned in the introduction, the purpose of this was to have one trial in which both partners experienced high mental load, one trial in which only the director experienced high mental load, one trial in which only the matcher experienced high mental load and one trial in which neither partner experienced high mental load. The order in which each dyad went through all four conditions (director high – matcher high, director high – matcher low, director low – matcher high and director low – matcher low) was counterbalanced across dyads.

After the end of the trials, the participants individually completed a sociodemographic questionnaire. They were also asked to say (a) whether they had been asked to memorise digits in each of the four experimental trials and (b) whether they believed that their partner had been asked to memorise digits in each of the four experimental trials. In line with our hypotheses, the purpose of this was to determine whether the participants were aware of the amount of mental load experienced by their partner (despite not being explicitly told that their partner had digits to memorise). All participants were then fully debriefed.

Data coding and experimental design

Once all data were collected, the dialogues between the participants were transcribed. All words produced as part of the task (including interjections, false starts, repetitions and hesitations; laughs were not included; we did not transcribe speech addressed to the experimenter, nor did we include the instructions provided by the experimenter to the participant). The transcriptions were then coded for feedback marker production. In order to determine which feedback markers should be included in our coding scheme, we included the French equivalent of all feedback markers usually examined in studies on this topic (see for instance Bangerter et al., 2004; Bangerter & Clark, 2003; Knudsen et al., 2020; Knutsen et al., 2018). This includes "yeah" (and similar forms such as "yeah" and "yep"), "okay", "right", "alright", "got it", "cool" and "really". In addition to this, other functionally similar feedback markers were also taken into account. The list of feedback markers (please note that the translation provided in brackets is necessarily approximate, as all feedback markers considered do not necessarily have a literal translation in English) taken into account in the current study was as follows: "ouais" or "oui" ("yeah" or "yes"), "ok" ("okay"), "d'accord" ("I agree"), "hmhm" ("mhm"), "bien" or "très bien" ("very well"), "ça marche" ("all good"), "ça roule" ("all good"), "je vois" ("I see"), "j'ai compris" ("I understand"), "ça va" ("all good"), "voilà" ("that's it"), "c'est bon" ("all good"), "c'est ça" ("that's it") and "j'y suis" ("I got there")[1]. The data from all trials (both successful and unsuccessful) were included in the analyses.

This coding scheme was then used to compute the three main dependent variables (DVs) in this study. Our first main DV was the number of feedback markers produced per speech turn, which was computed by adding up the number of feedback markers produced in each speech turn (regardless of which marker or markers this was). We then decided to focus on the two most frequent feedback markers in the corpus, namely "yeah" (and similar forms) and "okay" (see also Knutsen et al., 2018). Indeed, taken together, these two markers represented over 70% of all feedback markers produced in the experiment, and they were produced by all participants and all dyads (i.e., in each participant in the study said "yeah" and "okay" at least once). In contrast, all other markers represented 10% or less of all markers found in the corpus, and were not systematically produced by all participants in all dyads. Interestingly, "yeah" is typically used to signal horizontal transitions, whereas "okay" is typically used to signal vertical transitions (Bangerter & Clark 2003) Thus, our second main DV was the probability of producing "yeah" during a speech turn, and our third main DV was

[1] As detailed below, our analyses finally focused mainly on "yeah" and "okay", as these two markers represented over 70% of all markers found in the corpus. Although this was not the goal of this study, we checked whether "yeah" was mainly used as a horizontal marker, and "okay" as a vertical marker, as predicted by Bangerter & Clark (2003). In order to do this, we examined each marker produced in the corpus to determine (a) if the participants continued talking about the same piece of the puzzle after the marker had been produced (in which case this marker was coded as horizontal) or (b) if the participants switched to a different puzzle piece after the marker had been produced (in which case this marker was coded as vertical). In line with Bangerter and Clark, we found that in Experiment 1, "okay" was used as a vertical marker in 60.54% of cases, and "yeah" was used as a horizontal marker in 60.94% of cases; in Experiment 2, "okay" was used as a vertical marker in 63.39% of cases, and "yeah" was used as a horizontal marker in 62.08% of cases.
Because the participants were entirely free to perform the task in whichever way they wanted, as long as the matcher finally managed to solve the puzzles, we found that transitions within and across puzzle pieces were not always smooth (e.g., we found cases in which a participant seemed to signal their willingness to move on to the next puzzle piece by saying "okay", but the other participant continued to talk about the same puzzle piece because they actually needed more information to place it correctly). In other words, due to the spontaneity of the task, we found it difficult to determine whether the project markers should be coded based on the speaker's (inferred) intent, or based on the effect of the marker on the remainder of the dialogue. Future research should address the question of how mental load affects the function of project markers, using dialogue settings involving more explicit transitions across subprojects.

the probability of producing “okay” during a speech turn. This was done by coding each speech turn 1 if it included the marker of interest and 0 if it did not[2].

As is often done in studies on dialogue (e.g., Knutsen et al., 2018), additional DVs were also computed in this study: task success (i.e., whether each trial was performed successfully or not), the time taken to perform each trial[3], the number of speech turns produced per trial and the number of words produced per trial, and average utterance length. These variables are often measured in dialogue research to quantify collaborative effort, that is, the amount of work that the dyad puts into reaching mutual comprehension (e.g., Bangerter et al., 2020; Clark & Wilkes-Gibbs, 1986). In the current study, examining how these variables were affected by our mental load manipulation enabled us to determine whether mental load affects collaborative effort in addition to feedback production. Importantly, we decided *not* to include success on the digit task as one of our DVs in this study. Indeed, we believe that a participant not remembering the digits does not necessarily mean that they did not attempt to memorise them, nor that the mental load manipulation was unsuccessful in that case. For instance, a participant might have tried very hard to remember the digits during the interaction (and feel a very high load due to this throughout the entire dialogue) but still fail to remember the digits correctly. There were actually several instances of participants who actually mention this explicitly in the corpus from Experiment 2, in which the participants were not prevented from mentioning whether or not they had been giving digits to remember during the task.

The first two independent variables (IVs) were the director’s mental load (high or low) and the matcher’s mental load (high or low). Both of these IVs were within-participants

---

[2] This coding scheme implied that the number of data points included in each analysis was equal to the number of speech turns in the corpus.
[3] This measure was obviously less accurate than a reaction task measure, given the nature of the task used in this study.

variables. Participant role (director or matcher) was also included as a between-participant IV in some of the analyses (see detail below).

For clarity purposes, we will start by presenting the corpus as a whole. We will then present the results of the analysis on participants' awareness of each other's mental load. Then, we will turn to the analyses of the influence of our IVs on feedback production, providing a direct test of our hypotheses. Finally, we will report the results corresponding to the influence of our IVs on our measures of collaborative effort (task success, the time taken to perform each trial, the number of speech turns produced per trial and the number of words produced per trial, and average utterance length), thus describing how mental load affects the amount of effort put into the task by the participants.

Results

Corpus characterisation. The corpus gathered in Experiment 1 included 57,787 words, 41,127 (71.17%) of which were produced by the director and 16,660 (28.83%) were produced by the matcher. It included 6,530 speech turns. A total of 4,034 feedback markers were produced (see Table 1).

Table 1

List and Number of Feedback Markers Produced by the Participants in Experiment 1

Feedback marker	English translation	Number of occurrences in Experiment 1	Corresponding percentage
Ouais/oui	Yeah/yes	2,050	50.82%
Ok	Okay	853	21.15%
C'est bon	All good	428	10.61%
Hmhm	Mhm	204	5.06%
C'est ça	That's it	151	3.74%
D'accord	I agree	144	3.57%
Voilà	That's it	110	2.73%
Ca va	All good	26	0.64%

Je vois	I see	24	0.59%
Ca marche	All good	17	0.42%
Bien/très bien	Very well	13	0.32%
J'ai compris	I understand	13	0.32%
J'y suis	I got there	1	0.02%

Awareness of each other's mental load. Before conducting the main analyses on the effect of experienced and perceived mental load on dialogue and feedback production, we conducted a descriptive analysis of the final question which our participants were asked at the end of Experiment 1 (i.e., whether they had been given digits to remember in each trial and whether they believed that their partner had also been given digits to remember in each trial). We found that the participants' responses were almost always accurate when they were asked to assess their own mental load (i.e., directors provided a correct response 99.14% of the time and matchers provided a correct response 97.41% of the time), but accuracy dropped when the participants were asked to assess their partner's mental load (i.e., directors provided a correct response 43.10% of the time whereas matchers provided a correct response 52.59% of the time). This suggests that the participants remembered the mental load they had experienced very well but that they had difficulty estimating their partner's mental load. The detailed results are shown in Table 2. An inspection of hits and false alarms rates confirmed this interpretation.

Table 2

Participants' Beliefs about their Own Mental Load and their Partner's Mental Load in Each Trial

Directors' beliefs about their own mental load			
	Director responded that they were under high mental load	Director responded that they were under low mental load	Total

Director was under high mental load	57 (hits)	0 (misses)	57
Director was under low mental load	1 (false alarms)	58 (correct rejections)	59
Total	58	58	116

Matchers' beliefs about their own mental load

	Matcher responded that they were under high mental load	Matcher responded that they were under low mental load	Total
Matcher was under high mental load	56 (hits)	1 (misses)	57
Matcher was under low mental load	2 (false alarms)	57 (correct rejections)	59
Total	58	58	116

Directors' beliefs about the matchers' mental load

	Director responded that the matcher was under high mental load	Director responded that the matcher was under low mental load	Total
Matcher was under high mental load	32 (hits)	25 (misses)	57
Matcher was under low mental load	34 (false alarms)	25 (correct rejections)	59
Total	66	50	116

Matchers' beliefs about the directors' mental load

	Matcher responded that the director was under high mental load	Matcher responded that the director was under low mental load	Total
Director was under high mental load	30 (hits)	27 (misses)	57
Director was under low mental load	31 (false alarms)	28 (correct rejections)	59
Total	61	55	116

Note. The figures in this table correspond to the number of trials in each situation in terms of hits, misses, false alarms and correct rejections. For instance, the panel regarding the directors' beliefs about their own mental load should be read as follows: "When directors were asked about trials in which they had experienced a high level of mental load (57 trials in total), they responded that they had experienced a high level of mental load (i.e., they responded that they had been given digits to memorise) all 57 times. When they were asked about trials in which they had experienced a low level of mental load (59 trials in total), they responded that they had experienced a high level of mental load once and that they had experienced a low level of mental load 58 times."

Rationale of the inferential analyses. The data were analysed using linear and logistic mixed models in SAS OnDemand for Academics (GLIMMIX procedure). Linear models were used when the DV was numerical and logistic models were used when the DV was binary. Mixed models were chosen because our experiment involved three levels of analysis units: dyads, participants and items (i.e., the items were the tangram figures used). Mixed models include not only fixed effects, but also potentially random intercepts (which account for variability across analysis units) and random slopes (which account for variability in the units' sensitivity to the IVs included in the analyses). Following Barr et al. (2013), we started by including the maximal random effects structure justified by the design. However, not all random effects contribute to the model significantly. In such cases, the random effects which do not contribute to the model may be removed without affecting the model parameters (keeping them in the analysis would cause the model to fail to converge; Kiernan et al., 2012). Thus, the results reported hereafter correspond to the models from which such random effects were removed.

Due to experimenter error in counterbalancing, the number of trials in all four conditions was not exactly balanced. We obtained data from 28 dyads in the “director high – matcher high” condition, from 29 dyads in the “director high – matcher low” condition, from 29 dyads in the “director low – matcher high” condition and from 30 dyads in the “director low – matcher low” condition. We addressed this issue by correcting the degrees of freedom used in the analyses using Satterthwaite’s correction (Keselman et al., 1999; Satterthwaite, 1946).

All effects (main effects and interactions) were interpreted based on the associated *b* values (Jaccard, 2001). Analyses conducted on the data from Experiment 1 are listed in Table 3, in which detailed information on the random effects included in each analysis are also provided.

The data and the analysis code can be downloaded at https://osf.io/6uqpv/.

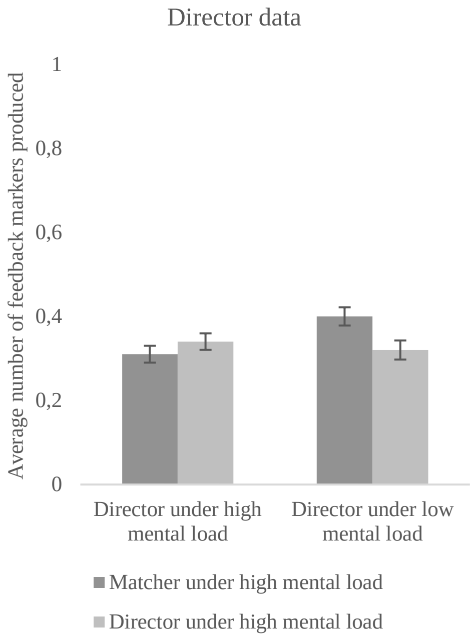
Table 3

Experiment 1 – Random Effects Structure used in Each Analysis

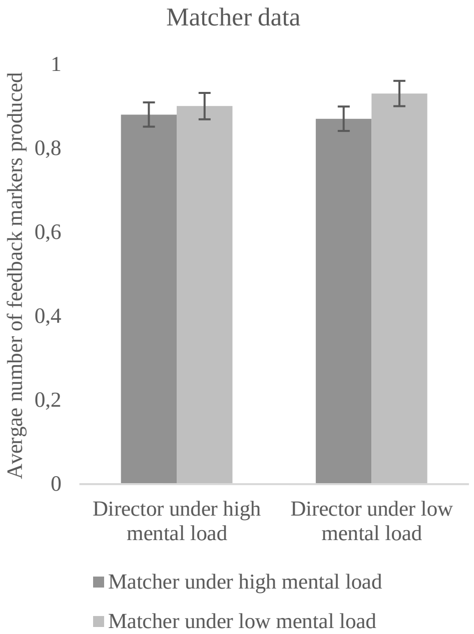
Analysis	Random intercepts	Random slopes
Analysis #1: Effect of participant role and director and matcher mental load on the number of feedback markers produced	No random intercepts included	By-dyad random slopes corresponding to participant role and director mental load; by-participant random slopes corresponding to director and matcher mental load and the number of words produced; by-item random slopes corresponding to participant role and the number of words produced
Analysis #2: Effect of participant role and director and matcher mental load on the probability of producing “yeah”	No random intercepts included	By-dyad random slopes corresponding to participant role, director mental load and the number of words produced; by-participant random slopes corresponding to matcher mental load and the number of words produced; by-item random slopes corresponding to participant role and the number of words produced

Analysis #3: Effect of participant role and director and matcher mental load on the probability of producing "okay"	By-participant random intercepts	By-dyad random slopes corresponding to participant role, matcher mental load and the number of words produced; by-participant random slopes corresponding to matcher mental load and the number of words produced; by-item random slopes corresponding to director mental load and the number of words produced
Analysis #4: Effect of director and matcher mental load on task success	By-item random intercepts	By-item random slopes corresponding to director mental load
Analysis #5: Effect of director and matcher mental load on the amount of time necessary to complete the task	By-dyad and by-item random intercepts	By-dyad random slopes corresponding to matcher mental load; by-item random slopes corresponding to director and matcher mental load
Analysis #6: Effect of director and matcher mental load on the number of speech turns produced	By-dyad and by-item random intercepts	By-dyad random slopes corresponding to matcher mental load; by-item random slopes corresponding to director and matcher mental load
Analysis #7: Effect of participant role and director and matcher mental load on the number of words produced	By-participant random intercepts	By-dyad random slopes corresponding to participant role and director and matcher mental load; by-participant random slopes corresponding to participant role and director and matcher mental load
Analysis #8: Effect of participant role and director and matcher mental load on utterance length	No random intercepts included	By-dyad random slopes corresponding to participant role; by-participant random slopes corresponding to director and matcher mental load; by-item random slopes corresponding to participant role

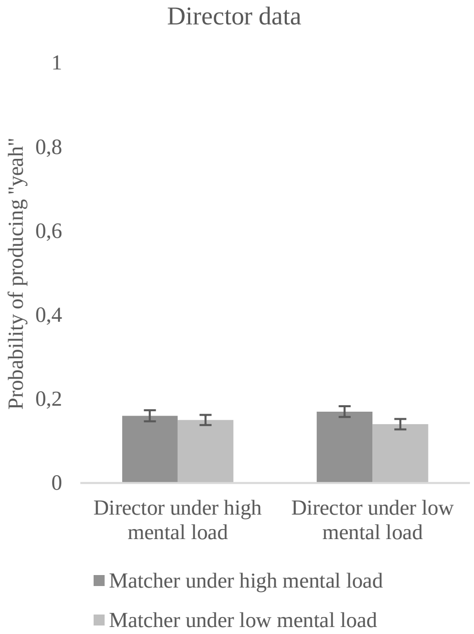
Effect of director and matcher mental load and participant role on the number of feedback markers produced (analyses 1-3). The data corresponding to analyses 1-3 are shown in Figure 3. Note that because the number of feedback markers produced could also depend on the total amount of words produced by the participants (e.g., dyads who tended to speak more were also more likely to produce more feedback markers), the number of words produced by the participants were included as a covariant in these three analyses.



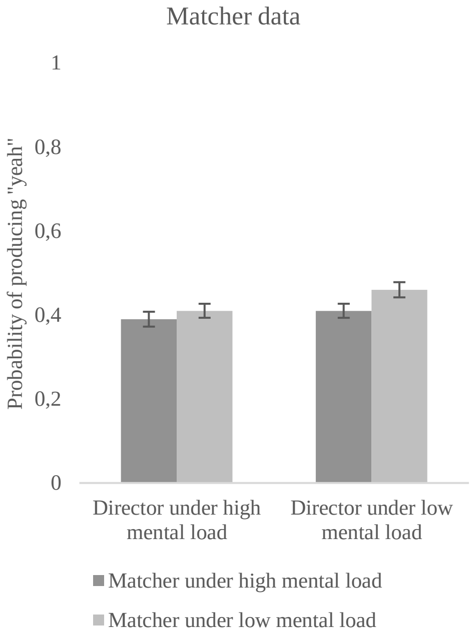
(a)



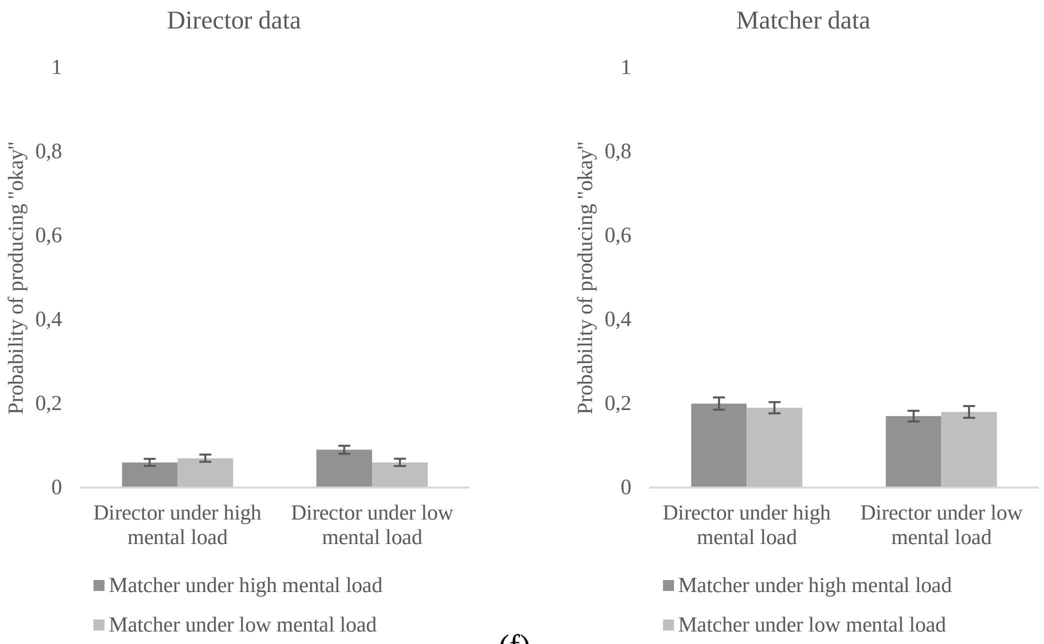
(b)



(c)



(d)



(e) (f)

Figure 3. Experiment 1 – Data for analyses 1-3. The director data are shown on the right and the matcher data are shown on the left. (a) and (b) represent the average number of feedback markers produced, (c) and (d) represent the probability of producing "yeah" and (e) and (f) represent the probability of producing "okay". The vertical bars represent the standard error.

The results of analysis 1 (number of feedback markers produced) are shown in Table 4. They revealed that the main effect of participant role was significant. The *b* value suggests that the director's utterances tended to contain fewer feedback markers that the matcher's utterances (see Figures 3a and 3b). There was also a significant main effect of the covariant. The *b* value suggests that the more the participants spoke, the *less* likely they were to produce feedback markers, which is somewhat surprising (one might have expected participants who spoke more to produce more feedback markers). All other effects failed to reach statistical significance.

Table 4

Experiment 1 – Results of Analysis #1

Effect	DFs	F	P	B	SE
Participant role	1, 43	107.23	<.001	-0.55	0.06
Director mental load	1, 25	0.96	.336	<-0.01	0.04
Participant role x director mental load	1, 17	0.16	.699	0.08	0.05
Matcher mental load	1, 47	<0.01	.991	0.01	0.04
Participant role x matcher mental load	1, 46	0.70	.406	0.03	0.06
Director mental load x matcher mental load	1, 2915	1.20	0.273	0.03	0.05
Participant role x director mental load x matcher mental load	1, 2090	3.49	.062	-0.14	0.07
Number of words produced	1, 33	15.25	<.001	-0.01	<0.01

The results of analysis 2 (probability of producing “yeah”) are shown in Table 5. They revealed that the main effect of participant role was significant. The *b* value suggests that the matcher’s utterances were more likely to contain the feedback markers “yeah” than the director’s utterances (see Figures 3c and 3d). There was also a significant main effect of the covariant. The *b* value suggests that the more the participants spoke, the *less* likely they were to produce the marker “yeah”. All other effects failed to reach statistical significance.

Table 5

Experiment 1 – Results of Analysis #2

Effect	DFs	F	p	B	SE	OR	$CI_{.95}$
Participant role	1, 46	35.91	<.001	1.08	0.24	3.18	2.16; 4.69
Director mental load	1, 33	0.57	.456	0.15	0.15	1.06	0.91; 1.23
Participant role x director mental load	1, 6521	1.34	.247	0.01	0.19	0.98	0.83 ; 1.17
Matcher mental load	1, 52	0.05	.825	0.09	0.17		
Participant role x matcher mental load	1, 52	0.69	.409	<-0.01	0.21		
Director mental load x matcher mental load	1, 4091	2.40	.122	-0.35	0.22		
Participant role x director mental load x matcher mental load	1, 6521	1.13	.287	0.29	0.27		

Number of words produced	1, 27	36.38	<.001	-0.09	0.01

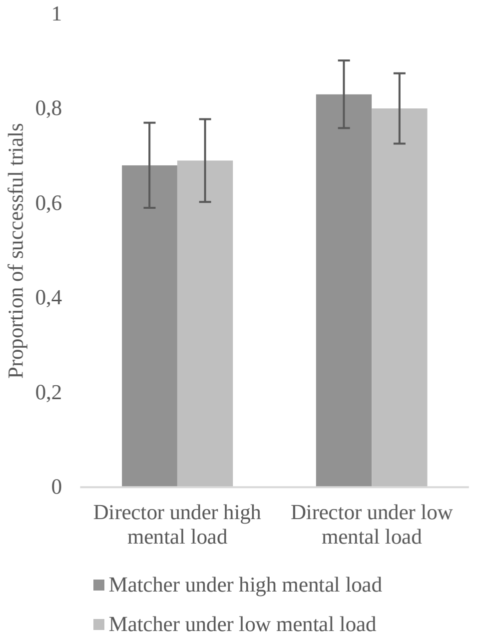
The results of analysis 3 (probability of producing “okay”) are shown in Table 6. They revealed that the main effect of participant role was significant. The *b* value suggests that the matcher’s utterances were more likely to contain the feedback marker “okay” than the director’s utterances (see Figures 3e and 3f). All other effects failed to reach statistical significance.

Table 6

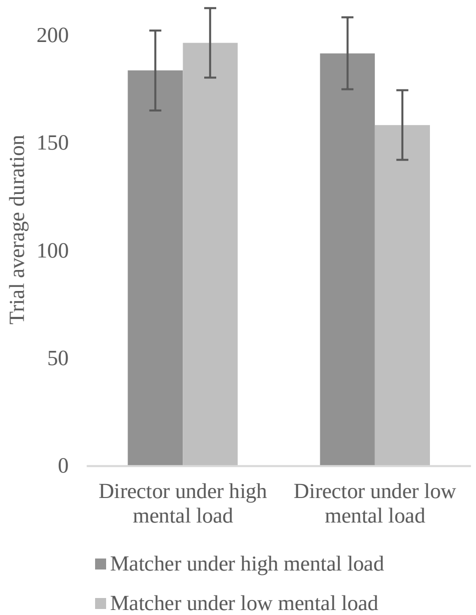
Experiment 1 – Results of Analysis #3

Effect	DFs	F	p	B	SE	OR	$CI_{.95}$
Participant role	1, 48	26.66	<.001	1.51	0.30	3.49	2.14; 5.67
Director mental load	1, 10	0.02	.898	0.36	0.22	1.02	0.79; 1.31
Participant role x director mental load	1, 3619	2.24	.135	-0.58	0.25		
Matcher mental load	1, 27	<0.01	.981	0.18	0.22	1.00	0.82; 1.23
Participant role x matcher mental load	1, 22	0.12	.730	-0.25	0.26		
Director mental load x matcher mental load	1, 1941	0.39	.533	-0.43	0.30		
Participant role x director mental load x matcher mental load	1, 6521	3.12	.077	0.63	0.36		
Number of words produced	1, 15	3.86	.068	-0.02	0.01		

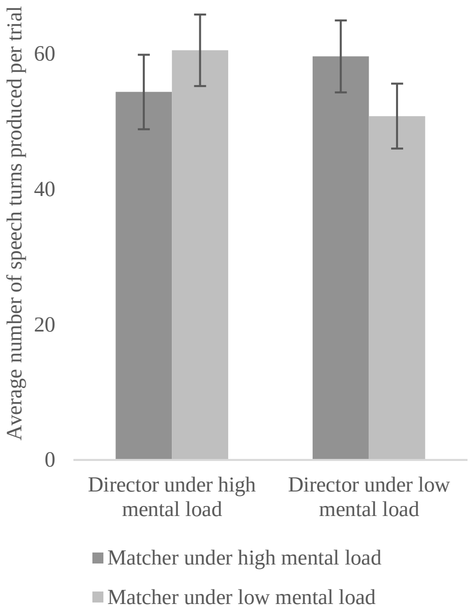
Effect of director and matcher mental load and participant role on collaborative effort (analyses 4-8). The data corresponding to analyses 4-8 are shown in Figure 4.


1
0,8
0,6
0,4
0,2
0
Proportion of successful trials
Director under high mental load
Director under low mental load
Matcher under high mental load
Matcher under low mental load

(a)


200
150
100
50
0
Trial average duration
Director under high mental load
Director under low mental load
Matcher under high mental load
Matcher under low mental load

(b)


60
40
20
0
Average number of speech turns produced per trial
Director under high mental load
Director under low mental load
Matcher under high mental load
Matcher under low mental load

(c)

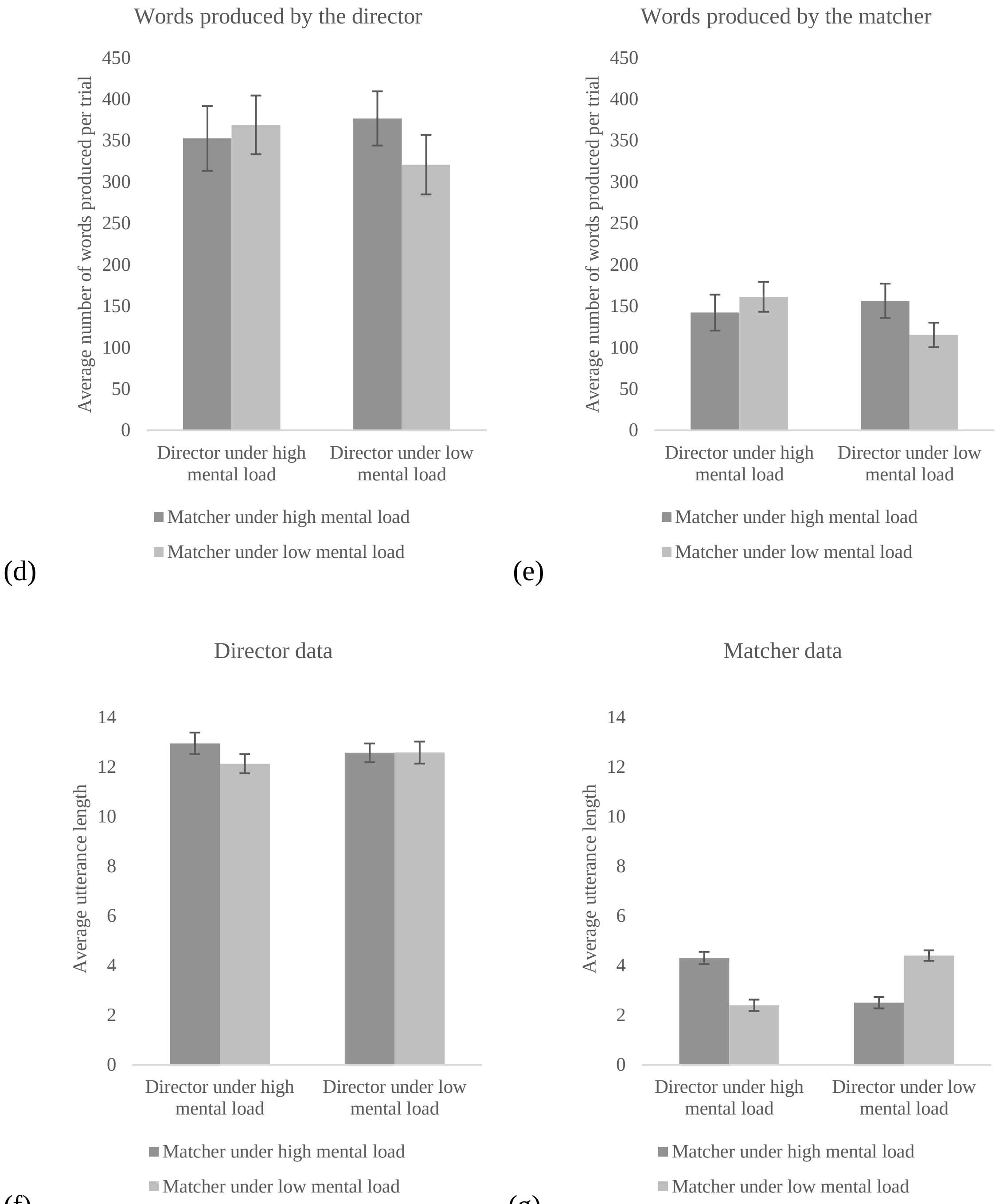


Figure 4. Experiment 1 – Data for analyses 4-8. (a) represents the proportion of successful trials, (b) represents average trial duration, (c) represents the average number of speech turns per trial, (d) and (e) represent the average number of words per trial and (f) and (g) represent average utterance length. The vertical bars represent the standard error.

Analysis 4 (success on the task) revealed that the success rate was overall quite high: the participants completed the task successfully on 75% of trials (see Figure 4a). The results of the analysis are reported in Table 7, which reveals that all effects failed to reach statistical significance.

Table 7

Experiment 1 – Results of Analysis #4

Effect	DFs	F	p	B	SE	OR	$CI_{.95}$
Director mental load	1, 8	1.49	.258	0.93	0.80	2.18	0.50; 9.45
Matcher mental load	1, 112	0.04	.845	0.06	0.61	0.91	0.37; 2.29
Director mental load x matcher mental load	1, 112	0.11	.745	-0.30	0.93		

The results of analysis 5 (time taken to complete a trial; see Figure 4b) are reported in Table 8 and revealed that all effects failed to reach statistical significance.

Table 8

Experiment 1 – Results of Analysis #5

Effect	DFs	F	P	B	SE
Director mental load	1, 4	1.95	.242	2.31	18.90
Matcher mental load	1, 7	0.20	.667	14.59	20.48
Director mental load x matcher mental load	1, 50	3.00	.089	-43.49	25.09

The results of analysis 6 (number of speech turns; see Figure 4c) are reported in Table 9 and revealed that although the main effect of director mental load and the main effect of matcher mental load failed to reach statistical significance, the interaction between these two factors was significant. The *b* value suggests that this was because when the director was under high mental load, the participants produced fewer speech turns when the matcher was under high mental load than when they were under low mental load, but that the opposite pattern was found when the director was under low mental load.

Table 9

Experiment 1 – Results of Analysis #6

Effect	DFs	F	P	B	SE
Director mental load	1, 2	0.86	.442	3.74	5.05
Matcher mental load	1, 6	0.01	.944	6.76	5.78
Director mental load x matcher mental load	1, 43	4.25	.040	-14.21	6.81

The results of analysis 7 (number of words; see Figures 4d and 4e) are shown in Table 10. They revealed that the main effect of participant role was significant. The *b* value suggests that the director produced more words than the matcher. There was also a significant director mental load x matcher mental load interaction. The *b* value suggests that this was because when the director was under high mental load, participants produced fewer words when the matcher was also under high mental load than when the matcher was under low mental load; however, the reverse pattern was found when the director was under low mental load. All other effects failed to reach statistical significance.

Table 10

Experiment 1 – Results of Analysis #7

Effect	DFs	F	p	B	SE
Participant role	1, 31	59.01	< .001	211.46	33.70
Director mental load	1, 5	0.45	.531	12.34	31.76
Participant role x director mental load	1, 90	0.01	.936	6.98	31.38
Matcher mental load	1, 10	0.09	.766	26.11	32.16
Participant role x matcher mental load	1, 91	0.13	.719	-2.70	31.37
Director mental load x matcher mental load	1, 116	7.53	.007	-59.70	32.35

Participant role x director mental load x matcher mental load	1, 92	0.06	.813	-10.45	44.13

The results of analysis 8 (utterance length; see Figures 4f and 4g) are shown in Table 11. They revealed that the main effect of participant role was significant. The *b* value suggests that the director's utterances tended to contain more words that the matcher's utterances. All other effects failed to reach statistical significance.

Table 11

Experiment 1 – Results of Analysis #8

Effect	DFs	F	P	B	SE
Participant role	1, 56	79.30	<.001	8.27	1.02
Director mental load	1, 42	1.41	.241	-0.11	0.58
Participant role x director mental load	1, 42	0.04	.839	-0.44	0.82
Matcher mental load	1, 40	0.47	0.496	0.25	0.53
Participant role x matcher mental load	1, 40	0.63	0.432	-0.76	0.75
Director mental load x matcher mental load	1, 3642	0.09	.764	-0.44	0.68
Participant role x director mental load x matcher mental load	1, 3642	0.39	.532	0.60	0.96

Discussion

The purpose of Experiment 1 was to determine whether the mental load experienced *and* perceived by both partners in a dyad influences feedback production during dialogue. Our main findings were that feedback marker production mainly depended on the role played in the dyad (director or matcher) whereas collaborative effort depended not only on the role played in the dyad, but also on the amount of load experienced and perceived by the participants.

Experiment 1 provided no evidence that feedback production depends on the amount of mental load experienced and/or perceived by dialogue partners. Indeed, the results revealed that feedback production mainly depended on participant role. Feedback markers were more likely to be produced by matchers than by directors; this was true regardless of whether feedback production was examined as a whole, or whether we focused specifically on markers which are representative of horizontal ("yeah") or vertical ("okay") transitions (Bangerter & Clark, 2003). One possible explanation (which is in line with prior studies in which workspace visibility was manipulated) is that because the director could not see the matcher's workspace, the matcher used a lot of feedback markers to indicate their progression in the task (Clark & Krych, 2004). The probability of producing feedback markers decreasing as the number of words increased is probably simply a by-product of this: matchers, who were more likely to produce feedback markers, also spoke less, explaining why participants who produced fewer words were also more likely to produce feedback markers. Moreover, the amount of collaborative effort put into the dialogue by the dyad varied as a function of participant role. In line with other research using the same task or a similar task (e.g., Knutsen et al., 2018, 2019), directors, who were required to provide precise instructions in order for their matcher to complete the task, spoke more and produced longer utterances than matchers.

Interestingly, however, the results confirmed that our mental load manipulation had a strong effect on the collaborative effort put into the dialogue by the dyad (Clark & Wilkes-Gibbs, 1986): the participants produced fewer speech turns and fewer words when they experienced the same level of mental load (i.e., both the director and the matcher were under high mental load or under low mental load) than when they experienced different levels of mental load (i.e., one of them was under high mental load while the other was under low mental load). Moreover, we found no evidence that mental load affected task success in our experiment. Dialogue is usually defined as more "efficient" when partners speak less (e.g., Bangerter et al., 2020; Clark & Wilkes-Gibbs, 1986). It thus seems that dialogue is less efficient when both partners experience different levels of mental load. Overall, partners experiencing different levels of mental load needed to produce more speech turns and words to reach the same rate of task success as when both they experienced the same level of mental load. This finding is all the more interesting that our preliminary analysis suggests that our participants were *not* explicitly aware of each other's mental load. We will return to this point in the General Discussion.

Taken together, these results presented some similarities with prior research (i.e., directors produced more words, longer utterances and fewer feedback markers than matchers) but they do not replicate the findings reported by Knutsen et al. (2018) on the impact of mental load on feedback production, nor do they enable us to validate our hypotheses: no conclusion can be drawn regarding the potential effect of mental load on feedback production. However, as mentioned in the introduction, one major difference between the current study and Knutsen et al.'s study is that our participants in Experiment 1 were not explicitly aware of each other's mental load. We thus conducted a second experiment in which participants were always told who, in the dyad, would receive digits to memorise. Our purpose was to determine whether people are capable of using this kind of information to modulate feedback production when it is provided explicitly.

Experiment 2

Participants

The participants were recruited in the same way as in Experiment 1. A total of 58 participants, who had not taken part in Experiment 1, took part in Experiment 2. Forty-four identified as female and 14 identified as men. Fourteen dyads included one female and one male and 15 dyads included only female participants. The participants' average age was 21.62 years (*SD* = 3.01).

Changes to the methodology used

The materials and the counterbalancing used in Experiment 2 were identical to those used in Experiment 1. The only difference between the two experiments was that in Experiment 2, the participants were explicitly told who would be given digits to memorise (this information was given out loud by the experimenter at the beginning of each trial). All other aspects of the procedure were identical to Experiment 1.

Results

Corpus characterisation. The corpus gathered in Experiment 2 included 56,268 words, 40,019 (71.12%) of which were produced by the director and 16,249 (28.88%) of which were produced by the matcher. It included 6,099 speech turns. A total of 3,737 feedback markers were produced (see Table 12). As in Experiment 1, the most frequent feedbacks markers were “yeah” and “okay”.

Table 12

List and Number of Feedback Markers Produced by the Participants in Experiment 2

Feedback marker	English translation	Number of occurrences in Experiment 2	Corresponding percentage
Ouais/oui	Yeah/yes	1842	49.29%
Ok	Okay	752	20.12%
C’est bon	All good	358	9.58%
Hmhm	Mhm	223	5.97%
D’accord	I agree	208	5.57%
C’est ça	That’s it	152	4.07%
Voilà	That’s it	90	2.41%
Je vois	I see	32	0.99%
Ca va	All good	30	0.80%
Ca marche	All good	24	0.64%
Bien/très bien	Very well	17	0.45%
J’ai compris	I understand	7	0.19%
J’y suis	I got there	2	0.05%

Inferential analyses were then carried out following the same rationale as in Experiment 1. The random effects included in all analyses are listed in Table 13.

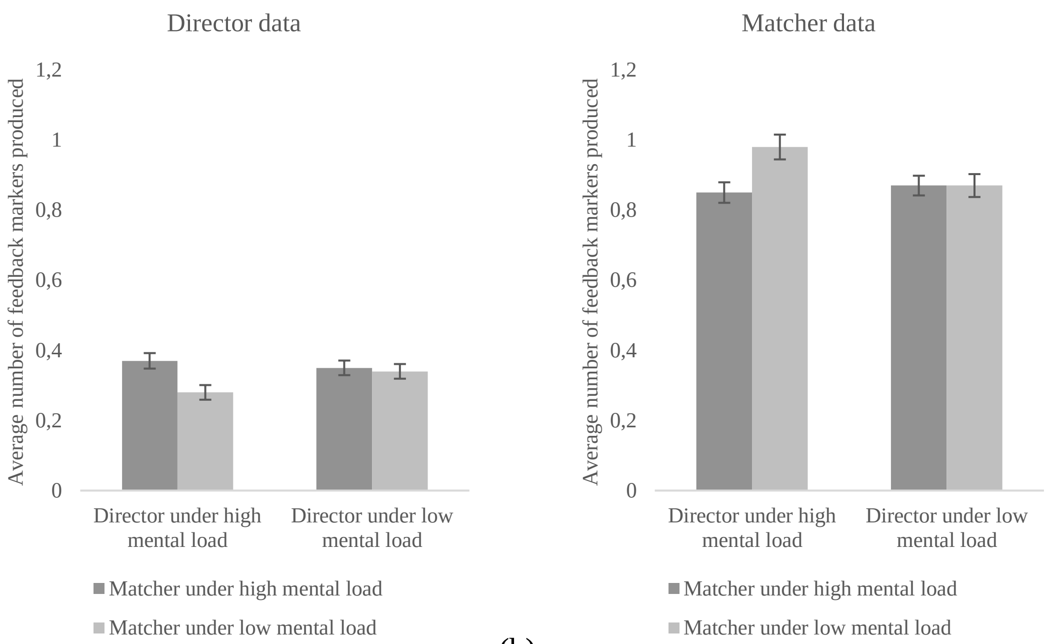
Table 13

Experiment 2 – Random Effects Structure used in Each Analysis

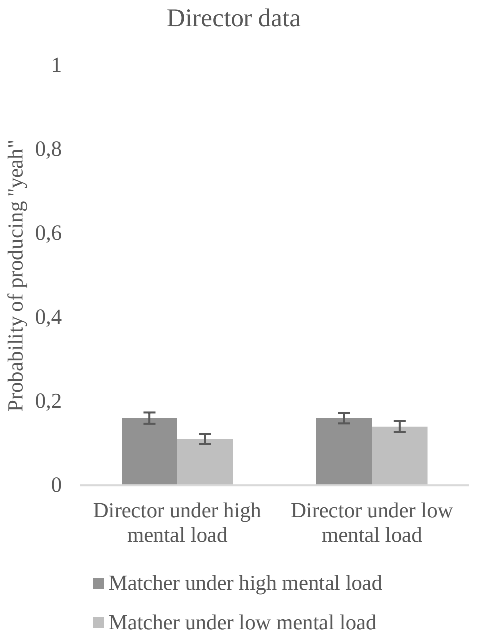
Analysis	Random intercepts	Random slopes
Analysis #1: Effect of participant role and director and matcher mental load on the number of feedback markers produced	By-participant random intercepts	By-dyad random slopes corresponding to participant role and the number of words produced; by-participant random slopes corresponding to director and matcher mental load and the number of words produced; by-item random slopes corresponding to participant role and the number of words produced
Analysis #2: Effect of participant role and director and matcher mental load on the probability of producing "yeah"	No random intercepts included	By-dyad random slopes corresponding to the number of words produced; by-participant random slopes corresponding to participant role, matcher mental load and the number of words produced; by-item random slopes corresponding to participant role
Analysis #3: Effect of participant role and director and matcher mental load on the probability of producing "okay"	By-participant random intercepts	By-dyad random slopes corresponding to director and matcher mental load
Analysis #4: Effect of director and matcher mental load on task success	By-dyad and by-item random intercepts	No random slopes included
Analysis #5: Effect of director and matcher mental load on the amount of time necessary to complete the task	By-dyad and by-item random intercepts	By-dyad random slopes corresponding to matcher mental load; by-item random slopes corresponding to director mental load
Analysis #6: Effect of director and matcher mental load on the	By-dyad and by-item random intercepts	By-dyad random slopes corresponding to director and matcher mental load; by-item random slopes corresponding to director mental load

number of speech turns produced		
Analysis #7: Effect of participant role and director and matcher mental load on the number of words produced	By-item random intercepts	By-dyad random slopes corresponding to participant role and to director and matcher mental load; by-item random slopes corresponding to participant role and to director and matcher mental load
Analysis #8: Effect of participant role and director and matcher mental load on utterance length	No random intercepts included	By-dyad random slopes corresponding to participant role; by-participant random slopes corresponding to director and matcher mental load; by-item random slopes corresponding to participant role and matcher mental load

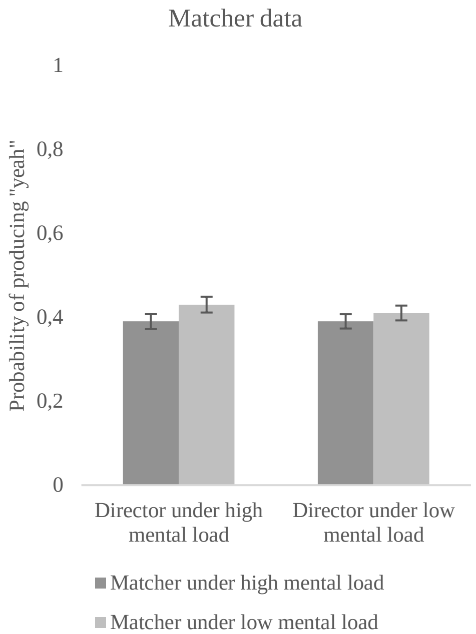
Effect of director and matcher mental load and participant role on the number of feedback markers produced (analyses 1-3). The data corresponding to analyses 1-3 are shown in Figure 5.



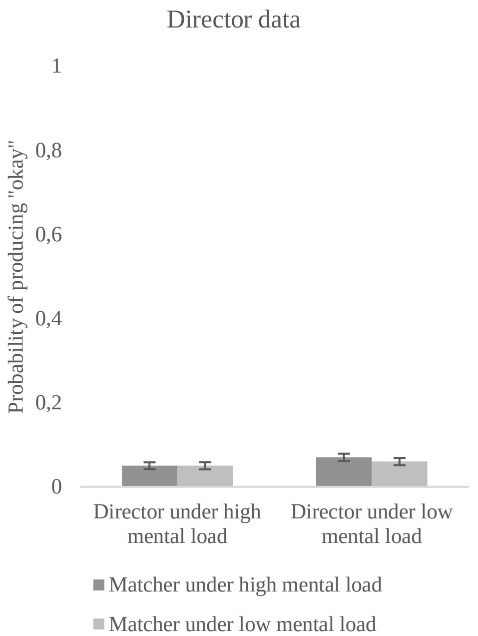
(a) (b)



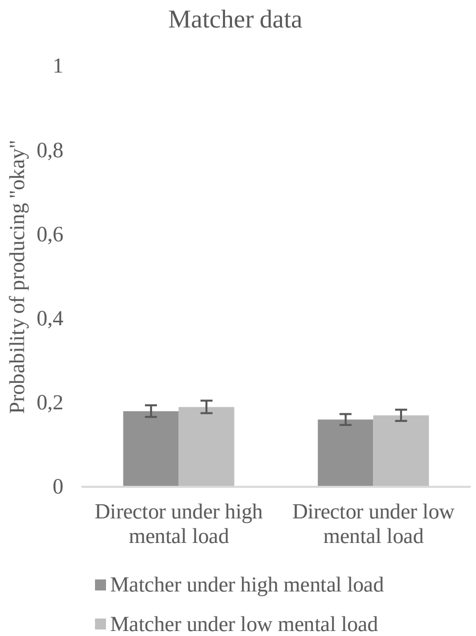
(c)



(d)



(e)



(f)

Figure 5. Experiment 2 – Data for analyses 1-3. The director data are shown on the right and the matcher data are shown on the left. (a) and (b) represent the average number of feedback markers produced, (c) and (d) represent the probability of producing “yeah” and (e) and (f) represent the probability of producing “okay”. The vertical bars represent the standard error.

The results of analysis 1 (number of feedback markers produced; see Figures 5a and 5b) are shown in Table 14. They revealed that the main effect of participant role was significant. The *b* value suggests that the director's utterances tended to contain fewer feedback markers that the matcher's utterances, as in Experiment 1. Unlike in Experiment 1, there was a significant participant role x matcher mental load interaction, which was qualified by a significant participant role x director mental load x matcher mental load interaction. An inspection of the *b* values suggests the following interpretation. The matcher produced more feedback markers than the director, especially when the matcher was under low mental load. This pattern was found mainly when the director was under high mental load. Finally, as in Experiment 1, there was a significant main effect of the covariate, with the associated *b* value suggesting that the more the participants spoke, the less likely they were to produce feedback markers. All other effects failed to reach statistical significance.

Table 14

Experiment 2 – Results of Analysis #1

Effect	DFs	F	P	B	SE
Participant role	1, 44	118.05	<.001	-0.47	0.06
Director mental load	1, 51	0.23	.634	0.03	0.04
Participant role x director mental load	1, 51	2.46	.123	-0.04	0.05
Matcher mental load	1, 44	0.02	.900	0.13	0.04
Participant role x matcher mental load	1, 44	8.23	.006	-0.22	0.06
Director mental load x matcher mental load	1, 2,105	1.05	.306	-0.14	0.05
Participant role x director mental load x matcher mental load	1, 2,150	7.46	.006	0.20	0.08
Number of words produced	1, 25	9.95	.004	-0.01	<0.01

The results of analysis 2 (probability of producing "yeah"; see Figures 5c and 5d) are shown in Table 15. They revealed that as in Experiment 1, the matcher's utterances were more likely to contain the feedback marker "yeah" than the director's utterances. Unlike in Experiment 1, there was also a significant participant role x matcher mental load interaction, which was qualified by a significant participant role x director mental load x matcher mental load interaction. An inspection of the *b* values suggests that this effect can be interpreted as in the previous analysis: the matcher produced more "yeahs" than the director, especially when the matcher was under low mental load. This pattern was found mainly when the director was under high mental load. Finally, as in Experiment 1, there was a significant main effect of the covariate, with the associated *b* value suggesting that the more the participants spoke, the less likely they were to produce the feedback marker "yeah". All other effects failed to reach statistical significance.

Table 15

Experiment 2 – Results of Analysis #2

Effect	DFs	F	P	B	SE	OR	$CI_{.95}$
Participant role	1, 43	40.01	<.001	0.90	0.22	3.17	2.19; 4.58
Director mental load	1, 6,090	0.03	.852	-0.06	0.15	1.01	0.88; 1.16
Participant role x director mental load	1, 6,090	2.76	.010	0.07	0.19		
Matcher mental load	1, 49	2.30	.136	-0.51	0.18	0.89	0.76 ; 1.04
Participant role x matcher mental load	1, 49	7.82	.007	0.73	0.22		
Director mental load x matcher mental load	1, 3,927	0.21	.650	0.37	0.23		
Participant role x director mental load x matcher mental load	1, 3,928	4.65	.031	-0.61	0.28		
Number of words produced	1, 26	39.33	<.001	-0.07	0.01		

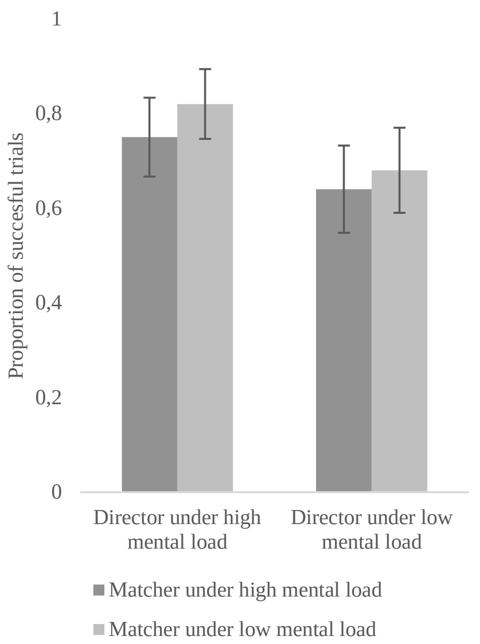
The results of analysis 3 (probability of producing "okay"; see Figures 5e and 5f) are shown in Table 16. Like in Experiment 1, they revealed that the main effect of participant role was significant. The *b* value suggests that the matcher's utterances were more likely to contain the feedback marker "okay" than the director's utterances. Unlike in Experiment 1, there was also a main effect of the covariate. The *b* values suggests that the more the participants spoke, the more likely they were to produce the feedback marker "okay". All other effects failed to reach statistical significance.

Table 16

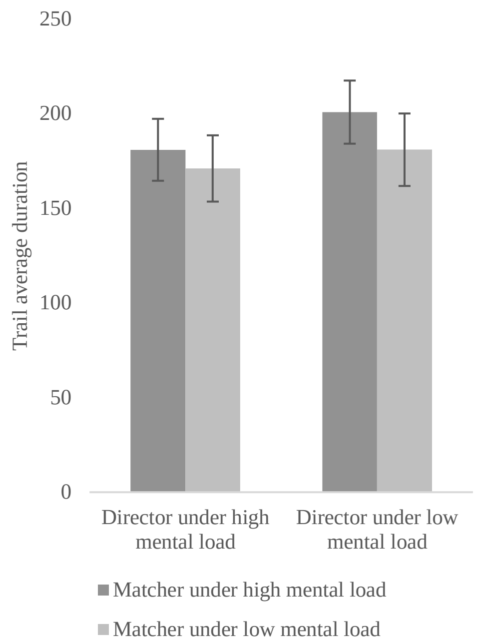
Experiment 2 – Results of Analysis #3

Effect	DFs	F	p	B	SE	OR	$CI_{.95}$
Participant role	1, 49	30.60	<.001	1.52	0.34	4.86	2.74; 8.62
Director mental load	1, 41	0.48	.491	0.18	0.24	1.08	0.87; 1.34
Participant role x director mental load	1, 1,282	1.25	.264	-0.23	0.27		
Matcher mental load	1, 28	0.09	.769	-0.22	0.27	0.96	0.75 ; 1.24
Participant role x matcher mental load	1, 1,521	2.89	.089	0.34	0.30		
Director mental load x matcher mental load	1, 6,090	0.01	.919	0.01	0.34		
Participant role x director mental load x matcher mental load	1, 6,090	<0.01	.959	0.02	0.39		
Number of words produced	1, 6,090	4.56	.033	0.01	<0.01		

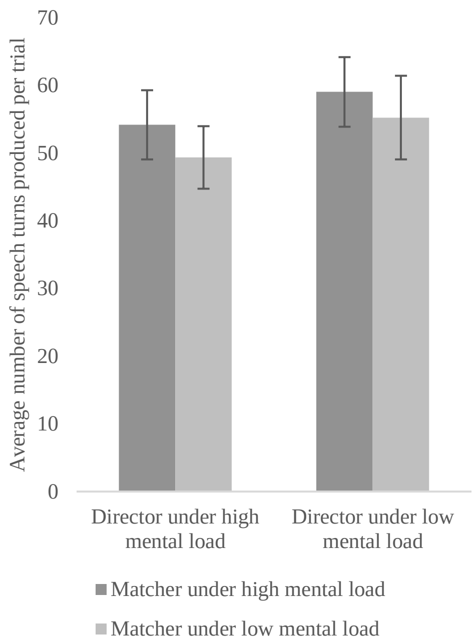
Effect of director and matcher mental load and participant role on collaborative effort (analyses 4-8). The data corresponding to analyses 4-8 are shown in Figure 6.


1
0,8
0,6
0,4
0,2
0
Proportion of succesful trials
Director under high mental load
Director under low mental load
Matcher under high mental load
Matcher under low mental load

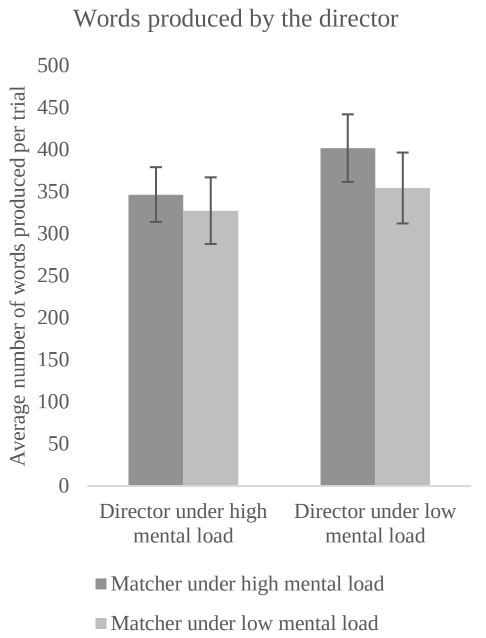
(a)


250
200
150
100
50
0
Trail average duration
Director under high mental load
Director under low mental load
Matcher under high mental load
Matcher under low mental load

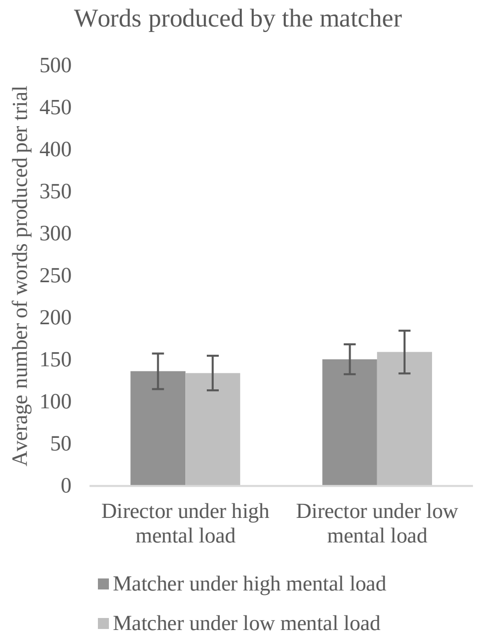
(b)


70
60
50
40
30
20
10
0
Average number of speech turns produced per trial
Director under high mental load
Director under low mental load
Matcher under high mental load
Matcher under low mental load

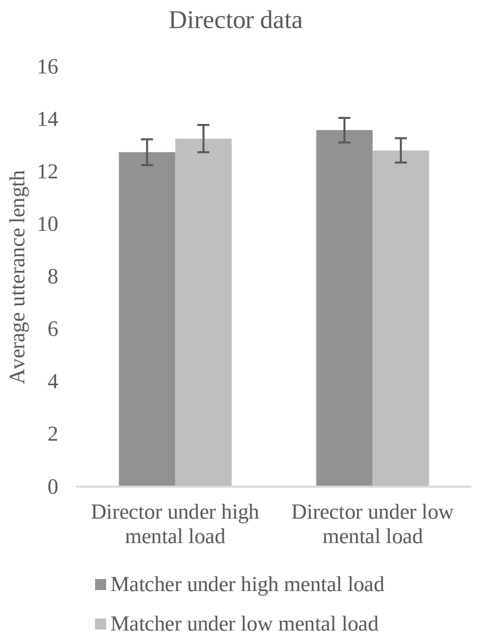
(c)



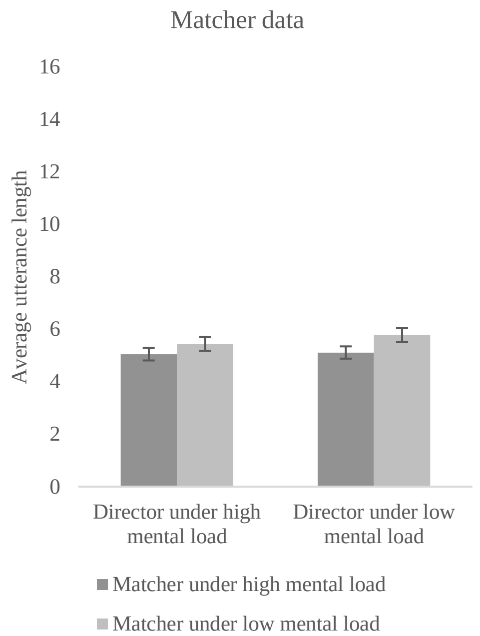
(d)



(e)



(f)



(g)

Figure 6. Experiment 2 – Data for analyses 4-8. (a) represents the proportion of successful trials, (b) represents average trial duration, (c) represents the average number of speech turns per trial, (d) and (e) represent the average number of words per trial and (f) and (g) represent average utterance length. The vertical bars represent the standard error.

Analysis 4 (success on the task; see Figure 6a) revealed that the success rate in Experiment 2 was quite high, just like in Experiment 1. The participants completed the task successfully on 72% of trials. The results of the analysis are reported in Table 17, which reveals that all effects failed to reach statistical significance.

Table 17

Experiment 2 – Results of Analysis #4

Effect	DFs	F	p	b	SE	OR	*CI*.95
Director mental load	1, 108	1.96	.164	-0.50	0.63	0.52	0.21; 1.31
Matcher mental load	1, 108	0.11	.741	0.31	0.70	1.17	0.46; 2.93
Director mental load x matcher mental load	1, 108	0.11	.740	-0.31	0.93		

The results of analysis 5 (time taken to complete a trial; see Figure 6b) are reported in Table 18 and revealed that just like in Experiment 1, all effects failed to reach statistical significance.

Table 18

Experiment 2 – Results of Analysis #5

Effect	DFs	F	P	B	SE
Director mental load	1, 8	0.68	.435	14.25	20.20
Matcher mental load	1, 35	< 0.01	.990	1.24	22.38
Director mental load x matcher mental load	1, 40	0.01	.909	-2.96	25.73

The results of analysis 6 (number of speech turns; see Figure 6c) are reported in Table 19 and revealed that unlike in Experiment 1, all effects failed to reach statistical significance.

Table 19

Experiment 2 – Results of Analysis #6

Effect	DFs	F	p	B	SE
Director mental load	1, 8	0.96	.357	3.88	6.30
Matcher mental load	1, 30	0.02	.884	-1.98	6.28
Director mental load x matcher mental load	1, 13	0.12	.731	2.46	7.01

The results of analysis 7 (number of words; see Figures 6d and 6e) are shown in Table 20. They revealed that the main effect of participant role was significant, as in Experiment 1. The *b* value suggests that the director produced more words than the matcher. However, unlike in Experiment 1, all other effects failed to reach statistical significance.

Table 20

Experiment 2 – Results of Analysis #7

Effect	DFs	F	p	B	SE
Participant role	1, 28	48.74	<.001	203.83	36.99
Director mental load	1, 5	0.96	.377	3.59	36.55
Participant role x director mental load	1, 81	1.10	.297	45.77	34.24
Matcher mental load	1, 16	<0.01	.980	3.17	35.60
Participant role x matcher mental load	1, 84	1.48	.228	-9.40	34.67
Director mental load x matcher mental load	1, 112	0.03	.868	24.90	36.18
Participant role x director mental load x matcher mental load	1, 82	0.71	.402	-40.82	48.40

The results of analysis 8 (utterance length; see Figures 6f and 6g) are shown in Table 21. They revealed that just as in Experiment 1, the main effect of participant role was significant. The *b* value suggests that the director's utterances contained more words than the matcher's utterances. All other effects failed to reach statistical significance.

Table 21

Experiment 2 – Results of Analysis #8

Effect	DFs	F	P	B	SE
Participant role	1, 52	38.03	<.001	9.11	1.47
Director mental load	1, 52	0.53	.470	0.46	0.53
Participant role x director mental load	1, 50	0.70	.407	-0.15	0.74
Matcher mental load	1, 14	0.12	.734	0.47	0.69
Participant role x matcher mental load	1, 43	2.65	.111	-0.90	0.92
Director mental load x matcher mental load	1, 952	0.43	.510	-0.03	0.74
Participant role x director mental load x matcher mental load	1, 657	0.38	.540	-0.64	1.04

Discussion

The results of both experiments are summarised in Table 22.

Table 22

Summary of the Results

Type of analysis	Analysis number	Results of Experiment 1	Results of Experiment 2	Were the results the same or different in both experiments?
Feedback marker analyses	Analysis #1 on the number of feedback markers produced	Main effect of role (Main effect of covariate)	Main effect of role Significant role x matcher mental load interaction Significant role x director x matcher mental load interaction (Main effect of covariante)	Partly (the effects of role are common to both experiments; the interactions were only found in Experiment 2)

	Analysis #2 on the probability of producing “yeah”	Main effect of role (Main effect of covariate)	Main effect of role Significant role x matcher mental load interaction Significant role x director x matcher mental load interaction (Main effect of covariate)	Partly (the effects of role are common to both experiments; the interactions were only found in Experiment 2)
	Analysis #3 on the probability of producing “okay”	Main effect of role	Main effect of role (Main effect of covariate)	Partly (the effect of the covariate was only found in Experiment 2)
Collaborative effort analyses	Analysis #4 on task success	No significant effect	No significant effect	Yes
	Analysis #5 on the time taken to complete the task	No significant effect	No significant effect	Yes
	Analysis #6 on the number of speech turns produced	Significant director x matcher mental load interaction	No significant effect	No (a significant director x matcher mental load interaction was found in Exp. 1)
	Analysis #7 on the number of words produced	Main effect of role Significant director x matcher mental load interaction	Main effect of role	Partly (in addition to the main effect of role, a significant director x matcher

				mental load interaction was found in Exp. 1)
	Analysis #8 on utterance length	Main effect of role	Main effect of role	Yes

The purpose of Experiment 2 was to examine the influence of mental load experienced *and* perceived by both partners in a dyad on feedback production, in a situation in which the participants were explicitly informed about the amount of mental load experienced by their partner. Our main findings were that feedback marker production depended both on the role played in the dyad and on experienced and perceived mental load, whereas collaborative effort only depended on the role played in the dyad.

The effect of mental load on the participants' speech was different in both experiments. Recall that in Experiment 1, the participants' mental load mainly affected our measures of collaborative effort (specifically, it affected the number of speech turns and words produced; these indicators mainly depended on whether or not the participants experienced similar levels of mental load). In Experiment 2, the participants' mental load affected the production of feedback markers: matchers produced more feedback markers than directors mainly when they (i.e., matchers) experienced low mental load *and* directors experienced high levels of mental load. This is consistent with our hypothesis that the effect of mental load is driven by one's partner's perceived level of mental load, in line with Knutsen et al.'s (2018) suggestion. In other words, we suggest that matchers in the current study produced more feedback markers to "help" directors under high levels of mental load. Moreover, this only occurred when matchers experienced low levels of mental load themselves, which is in line with our hypothesis that noticing that one's partner is experiencing high levels of mental load is only possible when one has enough mental resources to do so. This is consistent with previous findings which suggest that increased mental load prevents people from taking each other's dialogic needs into account (Rossnagel, 2000, 2004). Interestingly, we found no evidence that directors also helped matchers in the same way, as this effect was specific to matcher feedback production. This is also consistent with Knutsen et al.'s (2018) study, in which the matchers were more sensitive to the mental load manipulation that the directors. This could be explained by the fact that the directors' mental load was already high, as they were responsible for providing precise instructions to the matchers.

Importantly, part of the results obtained in Experiment 1 were replicated in Experiment 2. Specifically, just like in Experiment 1, we found no significant effect of role or of director or matcher mental load on task success or the time taken to complete the task. Directors produced more words, lengthier utterances and fewer feedback markers (including both "yeah" and "okay") than matchers. This confirms that the way in which each participant contributed to the task depending on their role was not affected by mental load awareness.

In sum, these findings show that matcher feedback production was affected both by the mental load they experienced themselves and the director's perceived mental load. This confirms that participants were able to use the information they had been provided about each other's mental load to modulate feedback production during the interaction. However, further analyses which distinguished between "yeah" (a marker which is mainly used to signal horizontal transitions) and "okay" (a marker which is mainly used to signal vertical transitions; Bangerter & Clark, 2003) suggested that the effect of mental load on feedback production was mainly driven by "yeahs" (experienced or perceived) was found on the feedback marker "okay". This is *not* consistent with Knutsen et al. (2018), who reported that the effect of mental load on matcher feedback production was mainly driven by vertical markers. This could be due to a number of differences between both studies, such as the type of mental load manipulation (time pressure in the 2018 study; memorising digits in the current study), the difficulty of the task (matchers could not see the completed figure in the 2018 study, whereas they could see it in the current study) or even the fact that both studies were conducted in different languages (British English in the 2018 study; French in the current study). Further studies should be conducted to explore the impact of the type of mental load manipulation and the difficulty of the task on collaboration in dialogue to better understand the functional link between horizontal and vertical markers and mental load.

General discussion

The purpose of this work was to determine how mental load affects feedback production in dialogue. Feedback markers play a central role in dialogue success insofar as they enable people to determine whether they have understood well enough for current purposes (Clark & Brennan, 1991) and to jointly manage project navigation throughout the interaction (Bangerter et al., 2004; Bangerter & Clark, 2003). Because dialogue partners may experience different levels of mental load as they interact, the participants' mental load was manipulated individually in the current study. This enabled us to distinguish between the influence of *experienced* mental load and *perceived* mental load (unlike in Knutsen et al.'s (2018) study). Our main hypothesis was that participants produce more feedback markers when their partners experience increased levels of mental load (in line with Knutsen et al., 2018). An additional hypothesis was that participants might fail to do so when they experience increased levels of mental load themselves, as mental load prevents people from collaborating efficiently (Rossnagel, 2000). Furthermore, examining whether the results from Experiment 1 and Experiment 2 were identical or not enabled us to determine whether people need to be made explicitly aware of each other's mental load for them to use this information during dialogue.

The results of Experiment 2 suggest that when people are explicitly told about each other's mental load, they attempt to use this information to adapt their production of feedback markers. Indeed, as mentioned previously, matchers produced more "yeahs" when the director was under high mental load, but only when the matcher had enough mental resources to do so (i.e., only when the matcher was under low mental load). Our interpretation is that this reflected the matchers' attempts to "help" the director by making grounding (Clark & Brennan, 1991) and dialogue navigation (Bangerter & Clark, 2003) more explicit. This interpretation is consistent not only with our hypotheses, but also with the more general idea that dialogue is a fundamentally collaborative activity (Clark, 1996; Clark & Wilkes-Gibbs, 1986). Our findings also highlight that mental load can affect dialogue in at least two different, opposite ways. On the one hand, mental load seems to boost collaboration, as perceived mental load causes people to attempt to help each more (see also Knutsen et al., 2018). On the other hand, experienced mental load may cause people to fail to identify their partners' dialogic needs (Rossnagel, 2000, 2004; see also Abel & Babel, 2017; Mattys et al., 2009, 2014; Mattys & Wiget, 2011, who showed that mental load impairs language processing outside dialogic settings). This helps explain why findings on this topic seem contradictory; we have suggested that carefully distinguishing between *experienced* and *perceived* mental load helps overcome this apparent contradiction.

According to the collaborative approach to dialogue (e.g., Clark, 1996, 2005; Clark & Brennan, 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986), dialogue partners follow the least collaborative effort principle as they interact. In other words, they attempt to reduce the total amount of effort spent by speakers and addressees during the conversation. In some cases, this principle implies that one of the participants puts more effort into the conversation than the other person (e.g., producing a well-adapted reference involves more effort from the speaker's point of view, but from the addressee's perspective, the effort required to understand and acknowledge the reference chosen is minimal; see also Schober, 1993, 1995). We suggest that the least collaborative effort principle can also help explain our findings. From the matcher’s perspective, producing more feedback markers when the director was perceived as potentially experiencing difficulty was probably cognitively costly, but doing so may have facilitated the director’s understanding, thus increasing the dyad’s overall chances of reaching mutual comprehension. The fact that experienced and perceived mental load mainly influenced the matchers’ behaviour is consistent with Knutsen et al.'s (2018) study. One possible explanation is that the task costs were particularly high for directors (who had to generate instructions for the matchers to follow), thus making it more difficult from them to identify and take into account their partners’ dialogic needs.

The results of Experiment 1 (in which the participants were provided no information about each other’s mental load) shed further light on our comprehension of dialogic feedback production. Our findings revealed no significant effect of mental load on feedback production, preventing us from validating or infirming our hypotheses. However, our mental load manipulation did have an influence on other aspects of dialogue. Indeed, we found that our participants were more “efficient” (i.e., they produced fewer words and speech turns) when their level of mental load was the same, regardless of whether they both experienced high levels or low levels of mental load. This finding is intriguing, as it implies that it is the correspondence between the partners’ state of mind (i.e., whether they both experience the same amount of mental load or not), rather than the actual amount of mental load experienced (i.e., whether their mental load was high or low), that determines dialogue efficiency. What is more, the participants’ responses to the questionnaires at the end of the study suggested that the participants were not explicitly aware of whether or not their partner was under high mental load, implying that such changes in the participants’ speech were not conscious or intentional.

One possible interpretation for this pattern of results is that two different mechanisms are at play in situations where participants are not explicitly aware of each other's mental load. On the one hand, participants (both) experiencing high levels of mental load may be limited in the number of words and speech turns they can produce, because language production is costly and that the amount of resources they may use to interact is limited. On the other hand, participants (both) experiencing low levels of mental load have enough resources to decrease their collaborative effort by individually producing fewer words and speech turns. Obviously, this interpretation is only speculative at this point, as we had not predicted this pattern of results when we designed this study. Future research should attempt to replicate and explain these findings. It is nonetheless interesting to note that when experienced by both members of a dyad, high levels and low levels of mental load led to the same result, that is, communication becomes more efficient.

In any event, it is important to highlight that interpreting our findings draws upon aspects of the collaborative approach to dialogue (Clark, 1996) on the one hand, and the egocentric approach to dialogue on the other (Barr & Keysar, 2002; Epley et al., 2004). Participants producing more or less feedback depending on their partner's mental load when explicitly provided this information (i.e., in Experiment 2) suggests that dialogue partners are opportunistic, in the sense that they use all information sources available during the interaction to produce partner-adapted utterances (see Clark & Krych, 2004), in line with the collaborative approach. However, the fact that information about one's partner's mental load had to be made explicit, and also that the participants were unable to adjust feedback production when they experienced increased levels of mental load themselves, implies that feedback production also depends on one's own state of mind during the interaction. Specifically, it depends on what one knows about their partner and the amount of mental load they are currently experiencing. This is in line with a more egocentric view of dialogue.

Taken together, then, our findings suggest that both attempts to collaborate and limitations inherent to the human cognitive system should be taken into account when trying to understand feedback production in dialogue.

Limitations of the study and plans for future research

A promising avenue for future research pertains to the nature of the mental load experienced by dialogue partners. Indeed, in the current study, the amount of mental load experienced depended on one of the features of the task (i.e., whether the participants were instructed to memorise digits, or not). However, in dialogue, a number of other factors are likely to increase speaker and listener mental load as well, some of them pertaining directly to the partners' features (e.g., adapting to a new dialogue partner might be more cognitively costly than adapting to an old friend) and some of them pertaining to the task at hand (e.g., talking about a new, complex topic, or talking in a foreign language, would be particularly costly). Sources of mental load may also be linguistic (e.g., as in the current study, in which participants had digits to memorise) or non-linguistic (e.g., participants might attempt to interact with someone while driving a car on a busy road). Additional research involving systematic variations in the nature of the cognitive load experienced by both partners would help us understand better how different aspects of the dialogue setting might affect the way in which people collaborate to reach mutual comprehension.

Another interesting question raised by our results is why perceived and experienced mental load affected the production of "yeah", but not "okay", in our study. One possibility would be to link this finding to the horizontal/vertical distinction described in the introduction (Bangerter & Clark, 2003). However, as highlighted in Footnote 1, an inspection of our two corpora suggested that not all yeahs were horizontal, and not all okays were vertical, in this study. Thus, more work would need to be done (possibly involving a different dialogue setting) to better understand how mental load influences vertical and horizontal transitions in dialogue.

One final question relates the prosody of the project markers produced by the participants in this kind of study. In the current study, we did not focus on the intonation used by the participants when they produced feedback markers. However, when listening to the recordings, we noticed that some feedback markers were produced with a rising intonation (e.g., “yeah(?)”). Our participants seemed to use these markers to elicit feedback from their dialogue partner (for a similar observation, see Knutsen et al., 2018). It would be interesting to examine whether dialogue partners under high mental load use this rising intonation more, thus causing their partners to provide more feedback.

To summarise and conclude, our findings suggest that people who are provided explicit information about each other’s mental load take this information into account to modulate feedback production during dialogue, although this may only occur if people have enough mental resources to do so. Not only do these findings document the influence of mental load on dialogue; they also highlight the importance of feedback production in mutual comprehension, as dialogue partners perceive feedback production as efficient enough to increase comprehension in cases where the other person might experience difficulties due to high levels of mental load. However, when our participants were provided no information about their partner’s mental load, other aspects of dialogue (such as the amount of effort put into the task) were affected by the amount of mental load experienced by both partners.

This raises the question of how close our two experiments are to what people experience in everyday conversational settings. It would be very uncommon for people to be explicitly given information about their partner’s mental load in everyday life (although the concept of mental load has received considerable public attention over the past few years; e.g., Dean et al., 2021). In addition, as highlighted above, different sources of mental load may be more or less “visible” to both dialogue partners, potentially affecting the likelihood that they would resort to this information to guide collaboration. However, people may use information about their prior knowledge and environment to make inferences about the amount of mental load experienced by other people. For instance, a passenger in a car may infer the driver’s mental load based on the amount of traffic surrounding the car or the presence or absence of adverse weather. Such prior knowledge was unavailable in our study, in which participants performed an unfamiliar task with an unknown partner whose face they could not see. Thus, the question of whether people routinely monitor the dialogue environment for this kind of cue (and more generally of how accurate people’s estimations of each other’s mental load is based on their knowledge and expertise) remains an open one and should be addressed in future studies involving more ecological settings.

References

Abel, J., & Babel, M. (2017). Cognitive load reduces perceived linguistic convergence between dyads. *Language and Speech*, *60*, 479-502. https://doi.org/10.1177/0023830916665652

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Éd.), *The Psychology of Learning and Motivation : Advances in Research and Theory* (Vol. 8, p. 47-89).

Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, *27*, 195-225. https://doi.org/10.1207/s15516709cog2702_3

Bangerter, A., Clark, H. H., & Katz, A. R. (2004). Navigating joint projects in telephone conversations. *Discourse Processes, 37*, 1-23. https://doi.org/10.1207/s15326950dp3701_1

Bangerter, A., Mayor, E., & Knutsen, D. (2020). Lexical entrainment without conceptual pacts ? Revisiting the matching task. *Journal of Memory and Language*, *114*, 104-129. https://doi.org/10.1016/j.jml.2020.104129

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, *46*, 391-418. https://doi.org/10.1006/jmla.2001.2815

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing : Keep it maximal. *Journal of Memory and Language*, *68*, 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Brennan, S. E., Galati, A., & Kuhlen, A. (2010). Two minds, one dialog : Coordinating speaking and understanding. In B. Ross (Éd.), *The Psychology of Learning and Motivation : Advances in Research and Theory* (Vol. 53, p. 301-344). Elsevier. http://dx.doi.org/10.1016/s0079-7421(10)53008-1

Brenner, M., Doherty, E. T., & Shipp, T. (1994). Speech measures indicating workload demand. *Aviation, Space, and Environmental Medicine*, *65*, 21-26.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Clark, H. H. (2005). Coordinating with each other in a material world. *Discourse Studies*, *7*, 507-525. https://doi.org/10.1177/1461445605054404

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Éds.), *Perspectives on socially shared cognition* (p. 127-149). American Psychological Association.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*, 62-81. https://doi.org/0.1016/j.jml.2003.08.004

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259-294. https://doi.org/10.1207/s15516709cog1302_7

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39. https://doi.org/10.1016/0010-0277(86)90010-7

Dean, L., Churchill, B., & Ruppanner, L. (2021). The mental load : Building a deeper theoretical understanding of how cognitive and emotional labor overload women and mothers. *Community, Work & Family*. https://doi.org/10.1080/13668803.2021.2002813

Della Salla, S., Foley, J. A., Beschin, N., Allerhand, M., & Logie, R. H. (2010). Assessing dual-task performance using a paper-and-pencil test : Normative data. *Archives of Clinical Psychology*, *25*, 410-419.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*, 327-339. https://doi.org/10.1037/0022-3514.87.3.327

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*, 931-939. https://doi.org/10.1177/0956797612436816

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8-11.

Gergle, D., Kraut, R. E., & Fussell, S. R. (2013). Using visual information for grounding and awareness in collaborative tasks. *Human–Computer Interaction*, *28*, 1-39. https://doi.org/10.1080/07370024.2012.678246

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index) : Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Éds.), *Human Mental Workload* (p. 139-183). North Holland Press.

Jaccard, J. (2001). *Interaction effects in logistic regression*. Sage.

Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1999). The analysis of repeated measurements : A comparison of mixed-model satterthwaite F tests and a nonpooled adjusted degrees of freedom multivariate test. *Communications in Statistics - Theory and Methods*, *28*, 2967-2999. https://doi.org/10.1080/03610929908832460

Kiernan, K., Tao, J., & Gibbs, P. (2012). *Tips and strategies for mixed modelling with SAS/STAT procedures*. 2012 SAS Global Forum, Orlando, FL.

Knudsen, B., Creemers, A., & Meyer, A. S. (2020). Forgotten little words : How backchannels and particles may facilitate speech planning in conversation? *Frontiers in Psychology*, *11*, 593671. https://doi.org/10.3389/fpsyg.2020.593671

Knutsen, D., & Brunellière, A. (2023, March 17). The effect of mental load on feedback production in dialogue. Retrieved from osf.io/6uqpv

Knutsen, D., Col, G., & Le Bigot, L. (2018). An investigation of the determinants of dialogue navigation in joint activities. *Applied Psycholinguistics*, *39*, 1345-1371. https://doi.org/10.1017/S0142716418000358

Knutsen, D., Mayor, E., & Bangerter, A. (2019). Procedural coordination in the matching task. *Collabra*, *5*, 3. https://doi.org/10.1525/collabra.188

Lysander, K., & Horton, W. S. (2012). Conversational Grounding in Younger and Older Adults : The Effect of Partner Visibility and Referent Abstractness in Task-Oriented Dialogue. *Discourse Processes*, *49*(1), 29-60. https://doi.org/10.1080/0163853X.2011.625547

Mattys, S. L., Barden, K., & Samuel, A. G. (2014). Extrinsic cognitive load impairs low-level speech perception. *Psychonomic Bulletin & Review*, *21*, 748-754. https://doi.org/10.3758/s13423-013-0544-7

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load : Dissociating energetic from informational factors. *Cognitive Psychology*, *59*, 203-243. https://doi.org/10.1016/j.cogpsych.2009.04.001

Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, *65*, 145-160. https://doi.org/10.1016/j.jml.2011.04.004

McInnes, F., & Attwater, D. (2004). Turn-taking and grounding in spoken telephone number transfers. *Speech Communication*, *43*, 205-223. https://doi.org/10.1016/j.specom.2004.04.001

Roque, A., & Traum, D. (2008). Degrees of grounding based on evidence of understanding. *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, 54-63.

Rossnagel, C. (2000). Cognitive load and perspective-taking : Applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology*, *30*, 429-445. https://doi.org/10.1002/(SICI)1099-0992(200005/06)30:3<429::AID-EJSP3>3.0.CO;2-V

Rossnagel, C. (2004). Lost in thought : Cognitive load and the processing of addressees' feedback in verbal communication. *Experimental Psychology*, *51*, 191-200. https://doi.org/10.1027/1618-3169.51.3.191

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, *2*, 110-114. https://doi.org/10.2307/3002019

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*, 1-24. https://doi.org/10.1016/0010-0277(93)90060-9

Schober, M. F. (1995). Speakers, addressees, and frames of reference : Whose effort is minimized in conversations about locations? *Discourse Processes*, *20*, 219-247. https://doi.org/10.1080/01638539509544939

Tolins, J., & Fox Tree, J. E. (2014). Addressee backchannels steer narrative development. *Journal of Pragmatics*, *70*, 152-164. https://doi.org/10.1016/j.pragma.2014.06.006

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*, 449-455. https://doi.org/10.1518/001872008X288394