Making oneself predictable in linguistic interactions

Jarosław R. Lelonkiewicz and Chiara Gambi

Scuola Internazionale Superiore di Studi Avanzati (SISSA)

Cardiff University

Word Count: 9583

Author Note

This is the final accepted manuscript of FULL reference to the paper.

Jarosław R. Lelonkiewicz, Cognitive Neuroscience Area, Scuola Internazionale Superiore di Studi Avanzati (SISSA); Chiara Gambi, School of Psychology, Cardiff University. The authors contributed equally.

This research was partially funded by a Research Support Grant from University of Edinburgh, School of Philosophy, Psychology and Language Sciences awarded to J.R.L; C.G was supported by a Leverhulme Trust Research Project Grant (RPG-2014-253) awarded to Hugh Rabagliati and Martin Pickering.

Raw data, analysis scripts, experiment scripts, and stimulus materials are available at Open Science Framework (https://osf.io/snk5g/).

Correspondence concerning this article should be addressed to Jarosław R. Lelonkiewicz, Cognitive Neuroscience Area, Scuola Internazionale Superiore di Studi Avanzati (SISSA), via Bonomea, 265, 34136 Trieste, Italy.

E-mail: jlelonki@sissa.it

Abstract

While language production is a highly demanding task, conversational partners are known to coordinate their turns with striking precision. Among the mechanisms that allow them to do so is listeners' ability to predict what the speaker will say, and thus to prepare their response in advance. But do speakers also play a role in facilitating coordination? We hypothesised that speakers contribute by using coordination smoothers – in particular by making their turns easier to predict. To test this, we asked participants to type definitions for common English words, either on their own (n = 26 individuals) or interacting with a partner (n = 18 pairs), and we measured the timing with which they produced the definitions. In a post-test, additional participants (n = 55) attempted to predict the final word of these definitions and rated them for quality. We found that interacting speakers initiated their turns with less variable delays than solo individuals. In contrast, our post-test measures suggested that jointly produced definitions were in fact of lower predictability and quality than those produced by individuals, but the analysis revealed these findings were likely confounded by task difficulty. We propose that the reduction in temporal variability observed for interacting speakers may facilitate prediction and thus act as a coordination smoother in linguistic interactions.

Keywords: coordination; prediction; dialogue; joint action; coordination smoothers

Making oneself predictable in linguistic interactions

One of the most striking observations about conversation is that people take turns with impressive fluency, leaving only small gaps between turns and very rarely overlapping with one another (Stivers et al., 2009). Recent evidence suggests that such fine linguistic coordination is supported by the listener's ability to predict how the current speaker's turn will unfold (Corps, Crossley, Gambi, & Pickering, 2018; Levinson, 2016; Lindsay, Gambi, & Rabagliati, 2019; Magyari, Bastiaansen, de Ruiter, & Levinson, 2014). In this paper, we investigate to what extent speakers may also facilitate coordination by making the timing and content of their turns easier to predict for their listeners.

The fluency of conversation is striking not only for the short latencies involved (median = 200 ms; Stivers et al., 2009), but more so because we know that the process of generating speech is computationally taxing and slow (Corps, Gambi, & Pickering, 2018; Indefrey & Levelt, 2004; Roelofs & Piai, 2011). This has led many researchers to ask the question of how people manage to achieve such a high level of coordination in joint language production. A growing number of studies are converging on the hypothesis that we achieve this by preparing our turns ahead of time (i.e., while our conversation partner is still speaking; Barthel, Meyer, & Levinson, 2017; Bögels, Magyari, & Levinson, 2015) and that we can do so because it is often possible to predict what our partner will say (Corps et al., 2018; Kutas, DeLong, & Smith, 2011; Pickering & Gambi, 2018). For example, Fjaellingsdal et al. (2020) asked a participant to build short sentences with a confederate, with the constraint that they had to alternate and could each contribute only one word at a time. The study showed that participant's turns were delayed after the confederate had produced an unexpected word, suggesting that participants attempted to predict the content of the confederate's turn (i.e., the confederate's next word) as listeners and that, when their prediction was disconfirmed, they took longer to generate their own turn.

But while there is good evidence that listeners prepare their turn by predicting what the speaker will say, this account places all the burden of coordination on the listener (i.e., the person currently listening and preparing to take the turn next). According to this account, the listener needs to undertake a fair amount of dual-tasking, as they must simultaneously comprehend the incoming turn, predict how the speaker's turn will unfold, and begin preparing their own turn. And although a number of studies have investigated the consequences of such multi-tasking and found evidence that it does take place (Barthel & Sauppe, 2019; Boiteau, Malone, Peters, & Almor, 2014; but see Sjerps & Meyer, 2015), we suggest that an undue emphasis on the listener's role risks detracting attention from the possible role of the speaker.

Indeed, research on joint action suggests that coordination often involves all interacting partners (Clark, 1996; Knoblich, Butterfill, & Sebanz, 2011; Pezzulo, Donnarumma, & Dindo, 2013; Richardson, Dale, & Shockley, 2008). Specifically, Vesper et al. (2010) proposed that joint action partners employ what they termed *coordination smoothers*, behaviours that facilitate the precise coordination of actions in space or time. One important example of this phenomenon is variability reduction: Reducing the variability in key movement parameters, such as speed, makes one's actions easier to predict, thus allowing the partner to plan and execute their own actions with greater precision. For example, Vesper et al. (2011) showed that when pairs of participants were instructed to coordinate either by pressing a button at the same time or in close temporal succession, they tended to reduce temporal variability in their actions, and the degree to which they did so predicted how well they coordinated with each other. Similarly, Vesper et al. (2016) showed that when pairs were given the goal of simultaneously landing on a target from different starting positions, they reduced variability in the duration of their reaching movements to facilitate coordination (at least when they could not observe each other's movement trajectories).

Moreover, theories of linguistic joint action (i.e., dialogue) also emphasise that coordination involves all conversation partners, rather than just listeners. For example, the interactive alignment theory of dialogue (Garrod & Pickering, 2009; Pickering & Garrod, 2004) argues that participants in a dialogue will progressively align their representations of the conversation at a number of linguistic levels (from phonetics to the conceptual model of the situation under discussion) thanks to priming from comprehension to production and vice versa. Importantly, it has been proposed that this process of convergence or alignment increases the similarity between partners, which in turn allows them to predict each other via a process of simulation (i.e., by working out what they themselves would say under the circumstances; Pickering & Gambi, 2018; Pickering & Garrod, 2013; see also Wilson & Knoblich, 2005 for related ideas in the joint action literature). In sum, alignment theories suggest that speakers linguistically align to one another and that one consequence of such alignment might be the facilitation of the interaction between the speakers.

But while there is indeed evidence that interlocutors converge on similar speech rates (Street, 1984; Schultz, O'Brien, Phillips, & McFarland, 2016; Webb, 1969) and that the more they align task-relevant linguistic expressions, the better they perform at a task that requires collective decisions (Fusaroli et al., 2012), it is less clear if partners in linguistic interactions may reduce variability as a way of smoothing coordination. Thus, in this study we ask whether variability reduction plays a role in linguistic interactions. In particular, we propose that, while listeners facilitate turn-taking by predicting and preparing ahead of time, speakers in turn aid their listeners by making themselves more predictable – specifically, by reducing variability in the timing and content of their turns.[1]

---

[1] Note that we do not mean to imply that speakers do this consciously or intentionally. Our study was not set up to address this question, but merely to establish whether speakers'

While we are not aware of any study that has explored the degree to which speakers make their turns more predictable in linguistic interactions, there is at least some indication that speakers, and not only listeners, play a role in ensuring the smoothness of turn-taking. According to so-called reactive theories of turn-taking (Duncan, 1972; Heldner & Edlund, 2010), the speaker signals when his or her turn is coming to an end (thus informing the listener they should get ready to take the floor). Several types of signals have been investigated, from gaze shifts towards the next speaker (e.g., Skantze, Hjalmarsson, & Oertel, 2014) to distinctive end-of-turn prosodic patterns (e.g., Bögels & Torreira, 2015). There is some evidence that such signals are spontaneously produced by speakers (Gravano & Hirschberg, 2011) and that do listeners react to them (e.g., Bögels & Torreira, 2015; Barthel et al., 2017). However, these signals typically occur too late in the speaker's turn to support prediction (hence, why these theories are termed "reactive"). Thus, while these signals could be viewed as coordination smoothers, this body of work cannot answer the question of whether current speakers facilitate coordination by making their turns more predictable to the listeners.

To our knowledge, a single study provides evidence that is relevant to answering this question (even though it was not actually set up for this purpose). Himberg et al. (2015) were the first to use a joint sentence production task where participants alternate producing words one at a time to jointly build a sentence; the study by Fjaellingsdal et al. (2020), mentioned above, later adapted this task for use with a participant and a confederate. In Himberg et al.'s (2015) study, two naive participants were simply asked to create a short narrative together (e.g., A: word 1, B: word 2, A: word 3, etc.). Participants were in different rooms and interacted either via an audio-

---

behaviour is indeed less variable in timing and content when they engage in a linguistic interaction.

only or an audio-and-video link (but since this manipulation made no difference to their findings, we will not discuss it further).

Himberg et al. (2015) did not analyse the content of the stories, and instead focused on the timing of participants' turns: They showed that participants entrained to each other's rhythm, producing their turns at a fairly constant lag with respect to their partner's last turn, despite the fact that the rhythms themselves were highly variable. However, this study lacked a control condition, so we cannot assess whether variability was reduced in the joint word production task compared to a task that did not require joint production. Further, since the content of the stories was not analysed, this study cannot address the question of whether participants made their turns more predictable in content.

We thus devised a novel task which is similar to Himberg et al. (2015) and Fjaellingsdal et al. (2020), but crucially included a control condition and allowed us to look at variability reduction in terms of both timing and content. We refer to this new task as the Word Chain Task, because it was loosely inspired by the popular 80's TV show Chain Reaction, where two participants produce a definition for a mystery word, contributing one word at a time. Similarly, in the joint version of our task, pairs of participants were given a mystery word to define. They were told they should produce a definition that would allow a third party to guess the mystery word. Participants sat in different booths and interacted via a chat-based interface (involving written text only). Crucially, the interface only allowed participants to contribute one word at a time, and further they were asked to constantly alternate (see example in Figure 1).

The control condition was the solo version of the task: In this solo version, a single participant was given the same mystery words and asked to produce a definition that would allow another person to guess it. Importantly, they typed their definitions in the same chat-based interface as used by participants in the joint condition and had to work with the same constraint of typing in one word at a time. However, they did not take turns with another person, so they produced all the words in the definition on their own (see Figure 1).

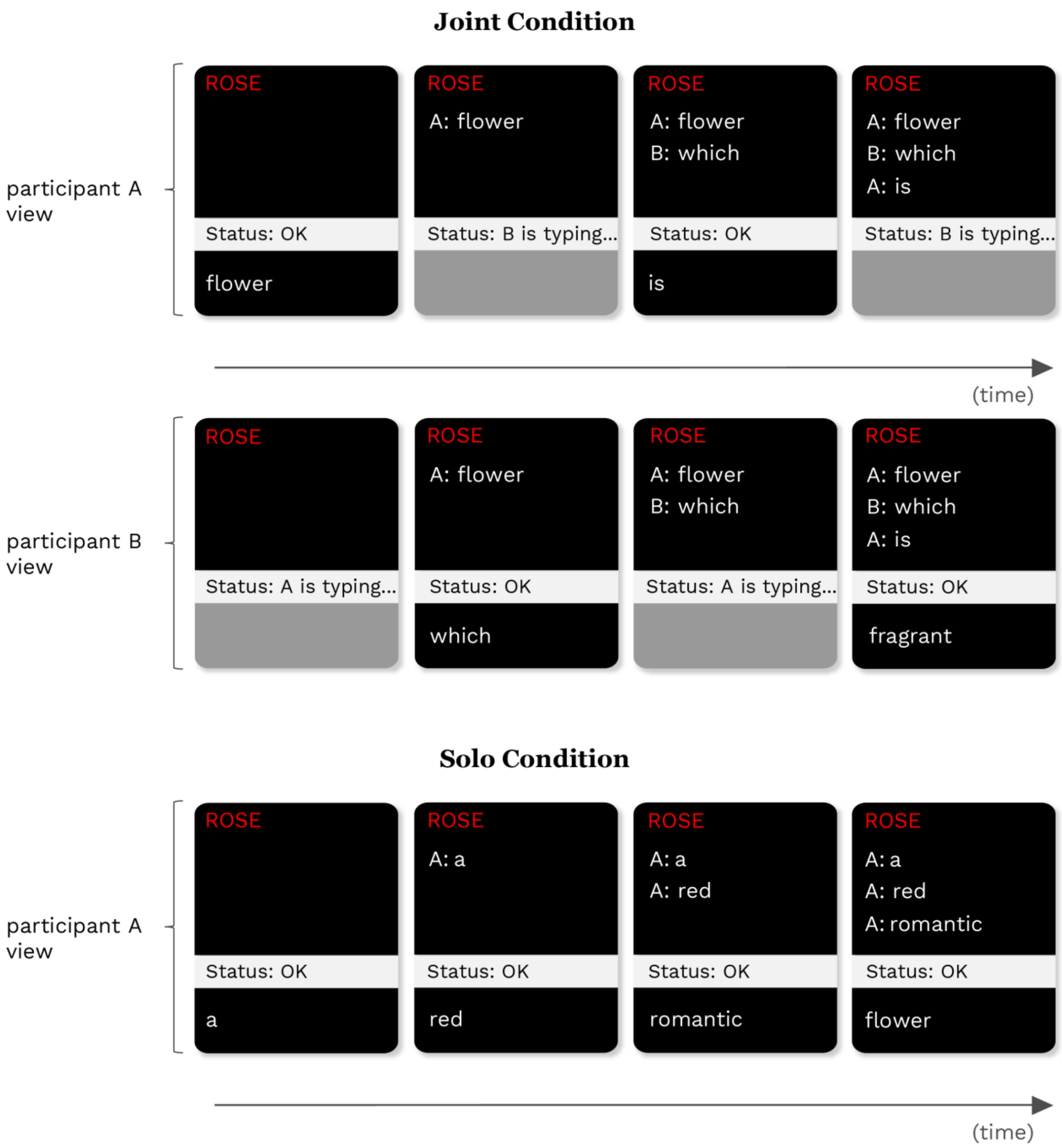

Figure 1. An illustration of the chat interface as seen by participants in the Word Chain Task. The upper panel depicts the joint condition (where two participants alternated producing a definition one word at a time), and the lower panel the solo condition (where a single participant produced a definition one word at a time). For simplicity, participants in this illustration are

represented by letters A/B, but in the experiment the software instead showed the nicknames as entered by the participants (e.g., Alex, Charlie).

By comparing the solo to the joint version of the task, we were able to investigate if the need to coordinate with another person would lead to the adoption of coordination smoothers, and specifically if the definitions produced by the interacting speakers were more predictable in timing and/or in content.

First, we asked whether participants in the joint version would reduce the temporal variability of their turns compared to participants in the solo version, thus making the timing of their turns more predictable. To answer this question, we assessed variability in both the duration of each turn (i.e., how long the participants spent typing each word) and the duration of the intervals between turns (i.e., the lag between the end of one turn and the beginning of the next). We expected variability in both of these timing measures to be lower in the joint version of the task, which required between-person coordination.

In addition, we asked whether participants in the joint version would reduce their variability also in terms of linguistic content - that is, whether they would produce turns whose content is more constrained so that it is easier to guess how they will unfold. The rationale here is that participants in the joint version of the task may support coordination with their partner by contributing words that are more likely to constrain “what comes next”, and thus facilitate the task of producing an appropriate continuation.

Currently, the best and most widely used measure of content predictability is the Cloze Task, which involves asking participants to read a sentence fragment and then contribute the first continuation that comes to mind (Taylor, 1953). The more participants converge on the same continuation, the more constraining the sentence, and the more predictable the continuation that participants converged on. Thus, we ran a post-test where a separate group of participants was presented with definitions that had been created in the Word Chain Task. The definitions were missing the final word (i.e., the final turn), and participants were asked to guess this word. We asked whether the predictability of definitions created by pairs in the joint version of the task would be greater than that of definitions created by solo participants.

In sum, we expected less temporal variability and greater content predictability in the joint task than in the solo task. However, the joint task was also in many respects more challenging than the solo task, and so we introduced an additional manipulation that could help us assess how task difficulty would affect our measures of variability and predictability. Why was the joint task more difficult than the solo task? Intuitively, this is because in the joint version of the task participants did not have access to each other's utterance plans, and needed to constantly revise their own utterance plans in response to each new turn produced by the other. Because this difference in difficulty may confound comparisons *between* the solo and the joint version of the task, we introduced an additional manipulation of task difficulty *within* each task, so that we could assess how increased task difficulty may have affected our measures of timing variability and content predictability independently from the comparison between solo and joint.

Our within-task manipulation of difficulty involved varying the ambiguity of mystery words. Specifically, half of the mystery words were ambiguous (i.e., they had at least two different meanings in English; e.g., BAT), while the other half were unambiguous (e.g., EGG). We reasoned that pairs in the joint condition would be more likely to experience difficulty when defining ambiguous words (compared to unambiguous words), because ambiguous words had a least two equiprobable meanings, increasing the likelihood that partners would select different meanings to define. Thus, partners would also be more likely to form misaligned utterance plans to begin with, and would have to work harder to resolve that misalignment, leading to more variable (in timing) and less predictable (in content) turns for ambiguous than unambiguous words in the joint condition. Thus, if task difficulty affects our measures of timing variability and content predictability, we would definitely expect an effect of ambiguity within the joint condition.

In contrast, it is less clear whether our ambiguity manipulation would affect participants in the solo version of the task: While individuals have of course access to their own utterance plans (so misalignment is not an issue), they may nevertheless notice the existence of additional meanings as they monitor their own utterances and attempt to include as many of these as possible in their definitions to meet the requirements of the task (i.e., to allow a third party to guess the mystery word more easily). If so, such additional self-monitoring and the increased demand on production processes may also result in more variable timing and less predictable content for ambiguous than unambiguous words in the solo condition. All in all, if task difficulty affects our measures of timing variability and content predictability, we would expect an effect of ambiguity within the joint condition, and perhaps also in the solo condition.

Finally, and as a further check of task difficulty, we used the post-test to assess the quality of the definitions produced by pairs and solo participants. If the joint task was indeed more challenging than the solo task, we would expect pairs to produce definitions of lower quality overall. Interestingly, however, previous research suggests that novel referring expressions are easier to understand by a third party when they have been produced as part of a dialogue, compared to when they have been produced by individuals in a monologue condition (Branigan, Catchpole, & Pickering, 2011; Tolins, Zeamer, & Fox Tree, 2018), and this may be because dialogue allows for multiple perspectives, but also requires these multiple perspectives to be negotiated down to a shared perspective. Thus, one might expect definitions produced by pairs to actually be of better quality than definitions produced by individuals.

Methods

Our study consisted of two separate experiments: The Word Chain Task and a post-test. In the Word Chain Task, participants typed definitions for common English words, either working with a partner (joint condition) or on their own (solo condition), and we measured the timing with which they produced the definitions. As a manipulation of task difficulty, participants in both conditions were asked to define an unambiguous word on half the trials, whereas on the other half they defined an ambiguous word. In the post-test, a further group of participants provided measures of content predictability and quality for the definitions produced in the Word Chain Task.

Our data, materials and commented analysis scripts are available at the Open Science Framework (OSF) page of this project: https://osf.io/snk5g/

Word Chain Task

Participants. We tested 18 pairs in the joint condition (11 female pairs, 2 male, 5 mixed), and 26 individuals in the solo condition (18 female, 8 male). Participants were native British English speakers, previously unacquainted, recruited from the University of Edinburgh student community, and were paid £4. The study was approved by the University of Edinburgh Psychology Research Ethics Committee.

Stimuli. Participants were asked to define 20 unambiguous words (e.g., EGG) and 20 ambiguous words (e.g., BAT; we selected words that had two or more meanings, but none of the meanings was much more likely than the others: dominant meaning probability ≤ .65 and ≥ .41). Ambiguous and unambiguous words were matched in terms of frequency and length (ambiguous: $M_{CELEX\ frequency}$ = 1107 per million; M_{length} = 1.4 syllables; unambiguous: $M_{CELEX\ frequency}$ = 1211 per million; M_{length} = 1.4 syllables; ps > .2). We also used three additional unambiguous words for the practice run stage of the task. All unambiguous words were nouns and the dominant meaning of ambiguous words was always a noun. See Table 1 for a list of words used in the Word Chain Task.

The ambiguous words were selected based on a pre-test where 48 participants (who did not take part in the Word Chain Task) estimated the frequency of every possible meaning for 37

English homonyms. On each trial, participants saw a word and the dictionary definitions for all its meanings (they were also allowed to add their own definitions) and they assigned percentages to indicate how often the word was used with the meaning described by each definition; based on these responses, we calculated the probability of each meaning for the ambiguous words used in the Word Chain Task. The pre-test was carried out using eDom software (Armstrong, Tokowicz, & Plaut, 2012).

Table 1. A list of ambiguous and unambiguous words used in the Word Chain Task. During the practice run of the task, participants defined three additional unambiguous words: ORACLE, HOLE, ROBOT.

ambiguous	unambiguous
BARK, BAT, BOARD, CALF,	BROTHER, CLOUD, CORNER, EGG,
CHANGE, CLUB, DEED, LINER,	HAIR, HERB, FLOUR, FOG,
MATCH, PANEL, POLICY, PRESENT,	FORT, KALE, LIMIT, MORNING,
PUPIL, RACE, RING, ROCK,	PUDDLE, RIVER, ROSE, SCANDAL,
RULER, SEAL, STRAW, TICK	SEED, SERVANT, SOAP, VASE

Apparatus. The set-up involved two client computers connected to a server. In the joint condition, computers formed a three-element network that allowed clients to communicate with each other via the server (i.e., client1-server-client2). In the solo condition, each client was connected to the server independently (i.e., client1-server, client2-server), and no communication between clients was allowed. Participants operated the client computers and typed the definitions into a customized chat software (DiET; Mills, 2014; Mills & Healey, retrieved from http://cogsci.eecs.qmul.ac.uk/diet/). The software's interface consisted of a window subdivided into three areas: (1) the main area where the mystery word was displayed (e.g., ROSE), along with the contributions to the current definition and the nickname of the participant who produced them, (2) the status box where the information about the other participant's typing activity was shown, and (3) the typing area. The status box displayed the following messages: In the joint condition, the message "[participant nickname] is typing.." was displayed if the other participant was currently typing, or otherwise "OK" was displayed; In the solo condition, where participants worked alone, the "OK" message was shown at all times. See Figure 1 for an illustration of the interface.

As participants were typing (i.e., adding a word to the definition), the software recorded all keystrokes along with their timestamps in server time. Once participants had finalized a turn (i.e., they finished typing a word and pressed ENTER), the server would summarize the associated keystrokes as a single word, and relayed it to the clients available in the network (e.g., r-o-n-[DELETE]-m-a-n-t-i-c-[ENTER] would be summarized as *romantic*; note that participants could make corrections while typing, but only the final resulting string was relayed to the clients). To ensure that joint participants would alternate when producing the definitions, the software enabled typing only for the participant who was expected to be contributing at the time (the typing area of the other participant would turn grey and the keyboard would be blocked; solo participants, however, were able to start typing a new word immediately after they had finished the previous one; for an illustration, see Figure 1).

Procedure. Two participants were recruited for each testing session. On arrival, they were seated in separate booths and both assigned to either joint or solo condition (condition assignment was determined by a virtual coin flip; https://www.random.org/coins/). They then received instructions for the task. In the joint condition, participants were asked to interact through the chat interface to produce definitions for common English words. They were told to take turns and to contribute only one word at a time. Participants in the solo condition were also asked to produce definitions word-by-word, but they worked separately and did not interact with each other (Figure 1).

In either condition, participants were instructed to complete the task as soon as possible and to construct their definitions according to three rules: The definitions could not mention the mystery word, could not be unintelligible or obscure, and had to involve at least one verb. To incentivize participants to follow these rules, we told them that another group of participants would later have to guess the mystery word for each definition. We also informed them that penalty points would be assigned for rule violation, and that the participant/pair of participants with the least penalty points would receive an additional £10 after the study was completed.

Next, participants entered their nicknames to log into the chat software, and carried out the Word Chain Task. In each trial, a mystery word appeared at the top of the chat window, signalling that participants could start producing the definition. In the solo condition, participants were able to contribute one word immediately after they had finished typing another. In the joint condition, however, typing was enabled for only one participant at a time, thus forcing participants to alternate (i.e., after participant A had finished typing a contribution, participant B was allowed to type theirs, followed by participant A again, etc.). To finish typing a word, participants pressed ENTER. To submit a complete definition, they typed "_" (underscore), after which the software wiped the chat window and the next trial started (in the joint condition, participants were allowed to type one extra contribution after their partner had signalled they wanted to submit the definition by typing "_"). Participants were allowed *ad libitum* time to produce a definition. Participants carried out three practice trials before the main task. The Word Chain Task took about 30 minutes to complete.[2]

Post-Test

Participants. We tested further 55 native British English speakers from the University of Edinburgh student community. Participants were paid £5. The study was approved by the University of Edinburgh Psychology Research Ethics Committee.

Stimuli, apparatus, and procedure. Participants were seated in front of a computer and explained they would see definitions that other people had produced for certain mystery words. They were told that, for each definition, they would be asked to guess what was the missing final word of the definition, guess the mystery word for which this definition was created, and finally rate how well the definition captured this mystery word (in this order). On each trial, participants saw a definition truncated by removing its final word, displayed centrally on the screen, one word at a time. They read the definition at their own pace, progressing from one word to another by pressing the SPACEBAR. After the definition had ended, a text box appeared and they were asked to type in the missing final word of the definition ("What is the missing final word of this sentence?"), followed by another question asking them to type in the mystery word for the definition ("The sentence you just saw was supposed to be a definition of a certain mystery word. Can you guess this mystery word?"). Finally, they rated the quality of the definition on a 1-5 Likert scale, ranging from extremely poor to extremely good ("The sentence you just saw was a definition of the mystery word [e.g., ROSE]. Please rate the quality of this definition by clicking on one of the numbers"), after which the next trial started. Each participant evaluated

[2] After the Word Chain Task, participants took part in a short pilot study investigating the link between linguistic cooperation and cooperation in decision making. This experiment was part of a different project and will not be reported here.

two different definitions produced for each of the mystery words, one randomly picked from the pool of joint, and another from the pool of solo definitions (e.g., each participant would see two definitions for ROSE, one produced by a pair, one by a solo participant); in total, each participant saw 80 definitions. The experiment was run in OpenSesame (v. 3.1.9; Mathôt, Schreij, & Theeuwes, 2012) and took approximately 30 minutes to complete.[3]

Results

Timing predictability

Measures and statistical analysis. Our first hypothesis was that joint participants would make the timing of their contributions more predictable than solo participants. Using the keystroke data from the Word Chain Task, we obtained two temporal measures of participants' language production: Typing time (i.e., total time spent typing a contribution to the definition) and inter-turn interval (i.e., time between the previous turn had ended and the beginning of the current turn; see Figure 2 for an illustration). Typing time was computed as the time between the onset of the first keystroke in a turn (e.g., *f* in *flower*) and when participants pressed ENTER to finish the turn (i.e., to signal they had finished typing the word). Inter-turn interval was the time between when ENTER was pressed and the onset of the first keystroke belonging to the following turn (e.g., *w* in *which*). These variables were computed in the same way in both solo and joint condition.

We operationalized timing predictability as the variability in the timing with which participants produced the definitions in the Word Chain Task (we reasoned that less variability

---

[3] Before completing the post-test, participants took part in a short (10 minutes) eye-tracking experiment, during which they listened to sentences while viewing pictures on the screen and where asked to point to the picture mentioned in each sentence. This experiment was part of a different project and will not be reported here.

means greater predictability). As an index of variability, we used the coefficient of variation (COV) defined as the standard deviation of a variable divided by the variable's mean (we chose the coefficient of variation because it controls for mean differences across conditions; see Abdi, 2010 for an introduction to this method; see Meyer, Bekkering, Haartsen, Stapel, & Hunnius, 2015 for a recent example of a study in which the COV was used as a measure of timing variability).

For either timing measure, we calculated the coefficient of variation (COV) for each definition produced by each author (either a solo participant or a joint pair), and entered these values in a linear mixed-effect model with Partner Condition (solo vs. joint) and Ambiguity (ambiguous vs. unambiguous) as predictors, and Mystery Word and Author as random effects. Predictors were contrast-coded and centered; the covariate was centered. All models reported in this article were computed using the *lme4* R package (v. 1.1-14; Bates, Maechler, Bolker, & Walker, 2019). The model had the form (in syntax of the *lmer* function): COV ~ 1 + Ambiguity * Partner Condition + Length + (1 + Ambiguity || Author) + (1 + Partner Condition | Mystery Word). Note that the model included random slopes for Partner Condition by Mystery Words (as the same words were given in either condition) and for Ambiguity by Author (as all participants/pairs were asked to define both ambiguous and unambiguous words). To help convergence, slopes and intercepts for Author were set as uncorrelated. Moreover, the model included Length as a covariate. Since the definitions produced by solo participants were on average longer than those produced by joint pairs (number of words in a definition in solo: M = 10.94, max = 44; in joint: M = 6.85, max = 23), and longer definitions were more variable (definition length was positively correlated with inter-turn interval COV: r = 0.76; and with typing times COV: r = 0.62), it was important to ascertain that differences in variability between partner conditions could not be fully accounted for by differences in definition length.

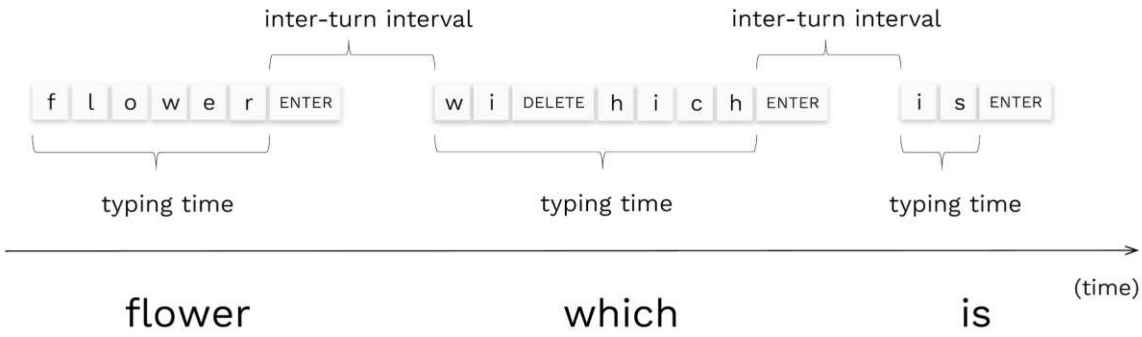

Figure 2. Temporal measures obtained from the Word Chain Task: Typing time was the time between the onset of the first keystroke in a turn (e.g., *f* in *flower*) and when participants pressed ENTER to finish typing the turn. Inter-turn interval was the time between when ENTER was pressed and the onset of the first keystroke belonging to the following turn (e.g., *w* in *which*).

Findings. The results are illustrated in Figure 3 and the model summaries are reported in Table 2. For convenience, results from gaussian models are reported with approximated p-values (*lmerTest* package v.3.1.0; Kutznetsova, Brockoff, & Christiansen, 2017). As expected, we found that inter-turn intervals were less variable in the joint than in the solo condition, suggesting that participants reduced the variability of the inter-turn intervals when working with a partner (t = -3.14, p = .003). In contrast, we observed that typing times were more variable in joint than in solo participants (t = 2.53, p = .015). However, this difference was driven by the increased variability in typing times for pairs defining ambiguous than unambiguous mystery words (Partner by Ambiguity interaction: t = 2.30, p = .026; see Figure 3), which suggests that joint participants spent longer producing the next word in a definition when their utterance plans were more likely to be misaligned. In contrast, ambiguity had no effect on the variability of inter-turn intervals (simple effect: t = -0.09, p = .928; Partner Condition by Ambiguity interaction: t = -1.02, p = .315), suggesting that joint participants were able to reduce variability in inter-turn intervals for both ambiguous and unambiguous words.

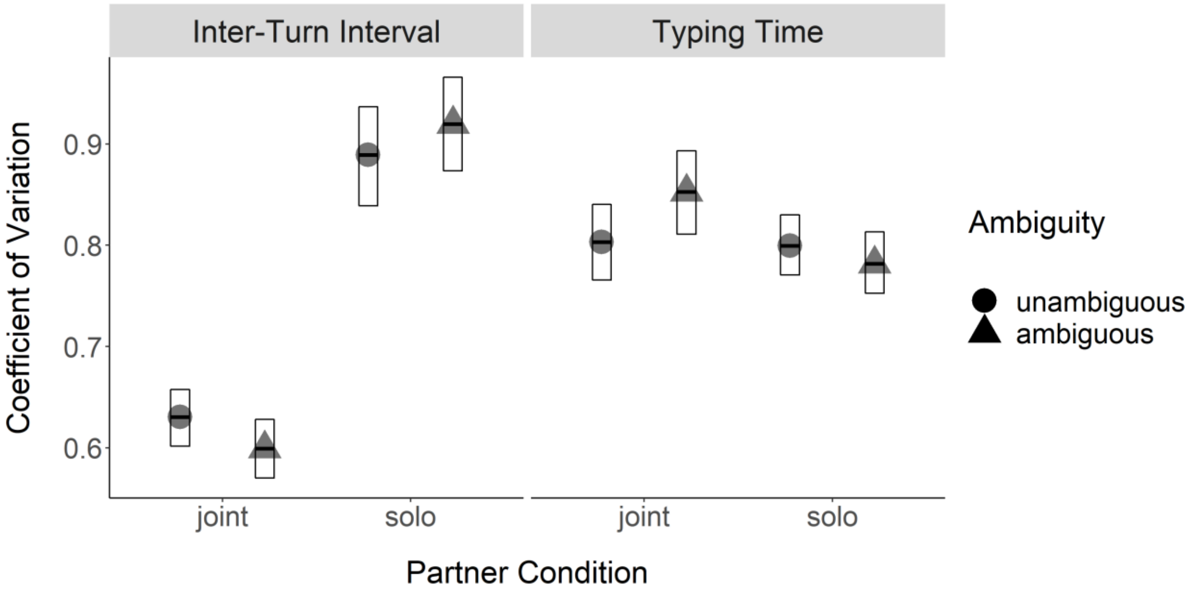

Figure 3. Timing predictability: Mean Coefficient of Variation for Inter-Turn Interval and Typing Time, presented by Partner Condition (joint vs. solo) and Ambiguity (unambiguous vs. ambiguous). The error bars represent 95% confidence intervals.

Table 2. Timing predictability: Summary of linear mixed-effect models for Inter-Turn Interval and Typing Time.

Inter-Turn Interval				
Fixed effect	*β*	*SE*	*t*	*p*
Length	0.034	< 0.01	14.22	< .001
Partner Condition	-0.150	0.05	-3.14	.003
Ambiguity	-0.002	0.03	-0.09	.983
Partner Condition:Ambiguity	-0.041	0.04	-1.02	.315
Random effect				Variance
Author				0.020
Mystery Word				0.004

Author:Ambiguity				0.002
Mystery Word:Partner Condition				< 0.001

Typing Time				
Fixed effect	*β*	*SE*	*t*	*p*
Length	0.016	< 0.01	8.00	< .001
Partner Condition	0.103	0.04	2.53	.015
Ambiguity	0.008	0.02	0.37	.717
Partner Condition:Ambiguity	0.075	0.03	2.30	.026
Random effect				Variance
Author				0.014
Mystery Word				0.002
Author:Ambiguity				< 0.001
Mystery Word:Partner Condition				< 0.001

Content predictability

Measures and statistical analysis. We hypothesized that pairs, as compared to solo participants, would produce definitions that are more predictable in terms of linguistic content. We used the post-test data to obtain two indices of content predictability. First, we defined predictability as the likelihood of correctly guessing the final word of a definition (when assessing the accuracy of participants' guesses, we allowed for an edit distance of one character to account for spelling mistakes). However, such a measure could be too conservative, as it assumes that a definition is predictable only when participants are able to guess the very same word as originally produced. This does not capture cases in which participants may have guessed lexical items that are not the same as the original completion, but are clearly semantically related. For example, when presented with a definition that reads "*flower which is fragrant* *and*___", some post-test participants completed it with the word "*pretty*" instead of the original "*beautiful*"; of course, "*pretty*" is strongly related to "*beautiful*", and we may wish our measure of content predictability to take this into account.

Thus, to provide a second, more sensitive measure of content predictability, we computed the average semantic similarity between the original final word of the definition and the completions provided by participants in the post-test. To capture semantic similarity we used Latent Semantic Analysis vector comparisons (Deerwester, Dumais, Furnas, Landauer, & Harsman, 1990): LSA determines the semantic similarity of words by calculating the extent to which they occur in the same contexts in a given corpus; it can range from 1 (words occur in identical contexts) to -1 (words never occur in the same context). Our LSA-based similarity measure was computed using package *LSAfun* in R (Günther, Dudschig, & Kaup, 2015), and word vectors were taken from the TASA semantic space for English (available at https://sites.google.com/site/fritzgntr/software-resources/semantic_spaces).

To test whether definitions produced by pairs were more predictable in content than those produced by solo participants, we ran (generalized) linear mixed models; when the dependent variable was the likelihood of correctly guessing the final word, we used a logistic link function (function *glmer* in R), whereas when the dependent variable was LSA-based predictability, we used a gaussian link function (function *lmer* in R). Both models included Ambiguity and Partner Condition as predictors (contrast-coded and centered), controlled for definition length (centered), and included random intercepts and slopes for Ambiguity and Partner Condition by Participant (as each post-test participant saw definitions produced for both ambiguous and unambiguous mystery words, and in joint and solo condition) and random intercepts for Definition. The models followed the syntax: predictability ~ 1 + Ambiguity * Partner Condition + Length + (1 + Ambiguity * Partner Condition | Participant) + (1 | Definition).

Findings. The results are illustrated in Figure 4, and model summaries can be found in Table 3 (logistic regression on the odds of guessing the final word) and Table 4 (gaussian regression on LSA-based predictability measure). Strikingly, we observed that joint participants produced definitions that were less predictable in terms of linguistic content, as compared to those produced by solo participants, and this was the case for both methods of measuring predictability (odds of guessing the final word: $z = 3.01$, $p = .003$; LSA-based predictability: $t = 2.23$, $p = .026$). We also found that predictability was lower for definitions produced for ambiguous than unambiguous mystery words ($z = -2.20$, $p = .028$; $t = -2.07$, $p = .039$). There was no interaction between Ambiguity and Partner Condition ($z = 0.98$, $p = .328$; $t = 0.64$, $p = .520$), suggesting that ambiguity led to lower content predictability both in the solo and in the joint condition.

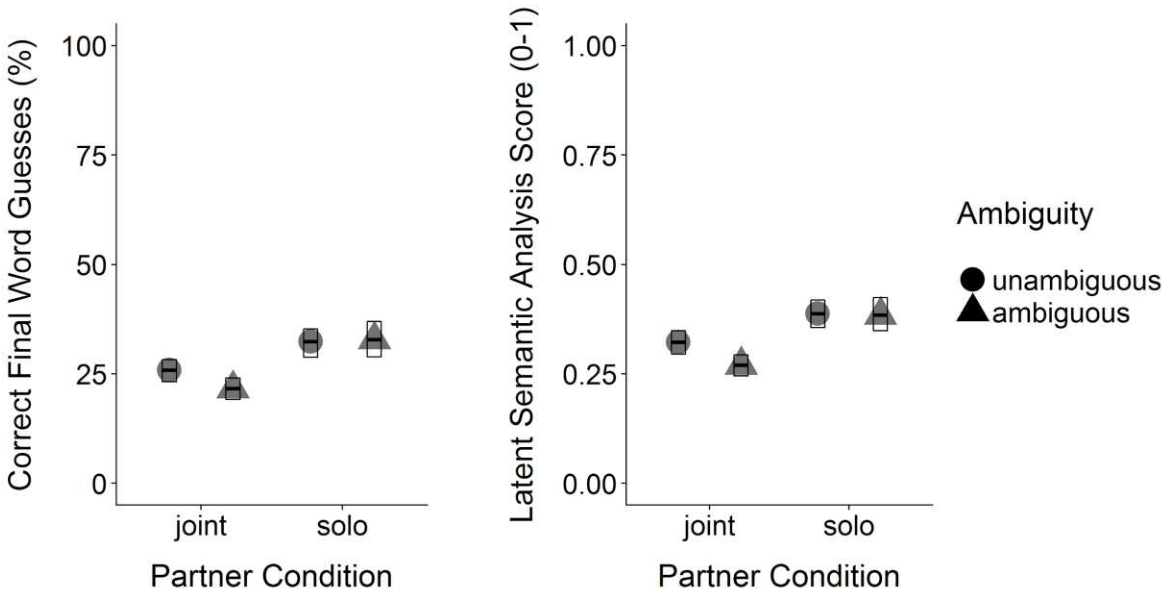

Figure 4. Content predictability: Mean percentage of correct final word guesses (left panel) and mean Latent Semantic Analysis (LSA) score for those guesses (right panel), by Partner

Condition (joint vs. solo) and Ambiguity (unambiguous vs. ambiguous). The error bars represent 95% confidence intervals.

Table 3. Content predictability: Summary of a generalized linear mixed-effect model (logistic regression) on odds of correctly guessing the missing final word of a definition.

Fixed effect	*β*	*SE*	*z*	*p*
Length	0.042	0.01	2.66	.008
Partner Condition	0.516	0.17	3.01	.003
Ambiguity	-0.347	0.16	-2.20	.028
Partner Condition:Ambiguity	0.307	0.31	0.98	.328
Random effect				Variance
Participant				0.044
Definition				4.387
Participant:Ambiguity				0.021
Participant:Partner Condition				0.051
Participant:Partner Condition:Ambiguity				0.041

Table 4. Content predictability: Summary of a linear mixed-effect model (gaussian regression) on LSA.

Fixed effect	*β*	*SE*	*t*	*p*
Length	0.008	< 0.01	4.25	< .001
Partner Condition	0.047	0.02	2.23	.026
Ambiguity	-0.040	0.02	-2.07	.039
Partner Condition:Ambiguity	0.025	0.04	0.64	.520
Random effect				Variance

Participant	0.001
Definition	0.080
Participant:Ambiguity	0.001
Participant:Partner Condition	0.001
Participant:Partner Condition:Ambiguity	0.103

Content quality

Measures and statistical analysis. Finally, we wanted to understand whether the definitions generated by pairs differed from those produced by individuals in terms of content quality. To this end, in the post-test we collected two measures of definition quality: An implicit measure based on participants' ability to guess the mystery word for the definition, and an explicit measure based on 1-5 Likert scale rating. We analysed these data with a generalized linear mixed model (for the odds of guessing the mystery word) and a linear mixed model (for ratings). The models had an analogous syntax as those used for the content predictability measures.

Findings. See Figure 5 for an illustration of results, and Table 5 (logistic regression on the odds of guessing the mystery word) and Table 6 (gaussian regression on average quality ratings) for model summaries. Interestingly, the pattern of results mimicked the one observed for content predictability: For both quality measures, we found that the definitions produced jointly had lower quality than definitions produced by solo participants (odds of guessing the mystery word: $z = 3.61$, $p < .001$; ratings: $t = 4.15$, $p < .001$), and that definitions for ambiguous words were of lower quality than definitions for unambiguous words ($z = -5.69$, $p < .001$; $t = -3.09$, $p = .002$). Again, the effects of Ambiguity and Partner Condition did not interact ($z = 1.07$, $p = .286$; $t = 1.82$, $p = .069$).

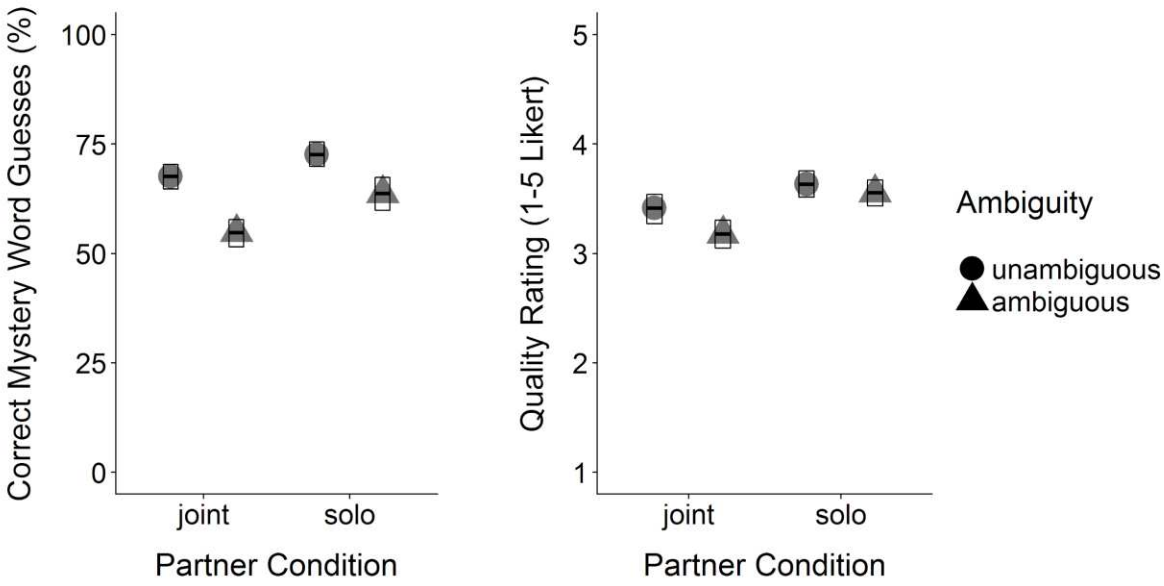

Figure 5. Content quality: Mean percentage of correct mystery word guesses (left panel) and explicit quality ratings (right panel), by Partner Condition (joint vs. solo) and Ambiguity (unambiguous vs. ambiguous). The error bars represent 95% confidence intervals.

Table 5. Content quality: Summary of a generalized linear mixed-effect model (logistic regression) on odds of correctly guessing the mystery word for the definition.

Fixed effect	*β*	*SE*	*z*	*p*
Length	-0.001	0.01	-0.12	.907
Partner Condition	0.513	0.14	3.61	< .001
Ambiguity	-0.756	0.13	-5.69	< .001
Partner Condition:Ambiguity	0.285	0.27	1.07	.286
Random effect				Variance
Participant				0.010
Definition				2.921
Participant:Ambiguity				0.008
Participant:Partner Condition				0.014

Participant:Partner Condition:Ambiguity	0.153

Table 6. Content quality: Summary of linear mixed-effect model (gaussian regression) on explicit quality ratings.

Fixed effect	*β*	*SE*	*t*	*p*
Length	0.011	< 0.01	2.12	.034
Partner Condition	0.234	0.06	4.15	< .001
Ambiguity	-0.158	0.05	-3.09	.002
Partner Condition:Ambiguity	0.183	0.10	1.82	.069
Random effect				**Variance**
Participant				0.159
Definition				0.502
Participant:Ambiguity				0.007
Participant:Partner Condition				0.015
Participant:Partner Condition:Ambiguity				0.004

Findings summary

In sum, our analysis showed that participants reduced the variability of the inter-turn intervals when working with a partner, as compared to working alone, and that they were able to do so even when the task was more difficult (i.e., when defining ambiguous mystery words). However, we found no evidence for a reduction in typing time variability. Further, with regards to the content of the definitions, we observed that joint participants produced definitions that were in fact less predictable and of lower quality than those produced by solo participants.

General Discussion

How do speakers and listeners make their conversations so well-coordinated? While there is mounting evidence that listeners contribute by predicting the speaker's turn, here we asked whether linguistic coordination may also be supported by speakers making their turns easier to predict. To address this question, we devised a novel task where participants typed definitions one word at a time, either alternating with a partner or on their own.

We found that inter-turn intervals (i.e., lags between consecutive words in a definition) were considerably less variable, and thus more predictable, in interacting participants than in solo individuals. Crucially, variability in inter-turn intervals was unaffected by our manipulation of task difficulty (i.e., by whether participants defined an ambiguous or unambiguous word), suggesting that the variability reduction is unlikely to have been confounded by differences in task difficulty between the joint and solo condition, and is instead more likely due to participants coordinating with one another. This finding is consistent with accounts suggesting that coordination in joint actions can be achieved through a mechanism of variability reduction (Vesper et al., 2010; 2013), and supportive of our proposition that interacting speakers make their turns more predictable to their partners, thus facilitating coordination.

Moreover, while other measures indicated that interacting participants were in fact less predictable than solo individuals, these same measures were also affected by our manipulation of task difficulty, suggesting that differences between interacting and solo participants in these measures may have also been driven by the greater difficulty associated with producing definitions jointly. Specifically, while definitions produced by interacting pairs were more variable in terms of turn durations (i.e., time spent typing a word) and less predictable in content, we also found that definitions were characterized by greater variability in turn duration and lower content predictability if they were produced for an ambiguous than unambiguous mystery word. Similarly, the quality of definitions produced by pairs was lower than that of definitions produced by individuals, and definitions of ambiguous words were of worse quality than those of unambiguous words.

One explanation of these findings is that differences in task difficulty impacted the participants' ability to plan their utterances. Utterance planning is the process of translating the concept behind the intended message (e.g., a to-be-defined mystery word) into a specific utterance or series of utterances (e.g., words in a definition). In most cases, such planning occurs in overlap with language production, with speakers generating and monitoring their plans while speaking (Meyer, 1996; Smith & Wheeldon, 2004). Manipulations that increase planning difficulty increase the proportion of errors, as well as affect the time it takes to produce an utterance (Smith & Wheeldon, 1999; Wagner, Jescheniak, & Schriefers, 2010). Thus, it is unsurprising that increased task difficulty influenced those of our measures that capture production outcomes (i.e., turn duration variability, content predictability, content quality). The fact that task difficulty did not influence the measure reflecting the lag between two consecutive production acts (i.e., inter-turn interval variability) suggests that participants may have engaged in utterance planning mainly during their turn, rather than between turns.

To explain the effect of ambiguity on turn duration variability, content predictability and content quality, we suggest that, in the context of our study, utterance planning was made more difficult when defining ambiguous mystery words. For solo participants, this additional difficulty may have been the result of participants' monitoring of their own definitions, that is, participants may have at times started a definition with the intention of capturing only one meaning of the mystery word, but later noticed the existence of additional meanings and decided to also include these to enhance the quality of their definitions. For instance, consider a participant who starts off by producing a definition for BAT - tool, but later realises that a person reading this definition could be misled into thinking it refers to CLUB - tool; to avoid the misunderstanding, a participant would have to revise their plan for the definition (and, consequently, for their upcoming contribution to the definition) to also include the meaning BAT - animal. Arguably, such cases would not be uncommon in our task - recall that participants were rewarded for producing a definition that would later allow others to correctly guess the mystery word.

Utterance planning would be similarly affected by having to produce a definition together with a partner. In the joint condition, participants were expected to generate a coherent definition that comprised also the words contributed by the partner, despite not knowing in advance what these words would be. Hence, their plan for the definition (and the individual words they wished to include in it) needed to take into account what the partner's contributions were likely to be at each turn - an additional challenge that was of course absent in the solo condition. Moreover, unless their partner behaved entirely predictably (i.e., in accordance with their plan), participants would sometimes observe their partner utter an unexpected word. In such instances, their plan for the definition had to be revised, which, again, required additional cognitive resources.

Hence, the need to more frequently revise utterance plans when defining an ambiguous mystery word or when interacting with a partner explains why in these conditions the task of producing definition was made more difficult, which in turn may have resulted in greater variability in typing times, and lower predictability and quality of the resulting definitions. Note that, since the task was unconstrained (i.e., participants generated their utterances freely), the level of planning difficulty likely varied throughout the course of each definition, and did so differently for each definition and participant/pair. This means that the level of difficulty may have not been uniformly higher when defining ambiguous words or in the joint condition, but one would still expect higher "peak" difficulty values and/or more frequent moments of "peak" difficulty, which would be reflected in higher timing variability in the more difficult conditions.

Finally, we suggest that for joint participants, the difficulty associated with having to incorporate the other's utterance plan into their own plan may have been further increased when defining an ambiguous mystery word. Since mystery words have at least two different meanings, and these meanings were chosen to be approximately equiprobable in the language, ambiguous words increased the likelihood of partners starting off with very different concepts and generating incompatible utterance plans. While this is an intriguing possibility, we should note that our data supported it only partially: Only our measure of turn duration variability showed an interaction between ambiguity and condition, such that the effect of ambiguity was larger in the joint than in the solo condition; the content predictability and definition quality measures instead suggested that ambiguity had a similar effect on both pairs and solo participants.

Since we cannot rule out the possibility that greater variability in turn duration and lower content predictability in the joint compared to the solo version of the Word Chain Task were due to increased task difficulty in the joint version, we instead focus on the finding that interacting speakers reduced the variability in the timing of inter-turn intervals. This finding is consistent with the proposition that joint actions are supported by a mechanism of variability reduction. In their minimal architecture for joint actions, Vesper et al. (2010) proposed that successful coordination requires agents to monitor and predict each other’s actions. This of course can pose a challenge, especially when access to information about the actions is limited. Thus, agents may at times make their actions easier to process by engaging in particular behaviors (or using certain objects), dubbed coordination smoothers. Indeed, several studies have demonstrated the use of coordination smoothers in joint actions - in particular, it has been shown that interacting agents aid coordination by constraining the spatial (Vesper et al., 2016) or temporal (Vesper et al., 2011; Vesper, Schmitz, Sebanz, & Knoblich, 2013) variability of their actions. Critically, evidence for the use of coordination smoothers comes predominantly from tasks involving motor movements. By showing that variability reduction may occur also in conversations, our study supports the proposition that these behaviors are engaged across a wide variety of joint actions (Vesper et al., 2010).

Interestingly, the possibility that coordination smoothers are engaged in conversations is convergent with several lines of research. First, it is compatible with prediction-centered models of language processing, which are built on the assumption that people routinely predict the next word they will hear or read (e.g., Altmann & Kamide, 1999; Federmeier & Kutas, 1999; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). These models argue that, in a conversational context, both speakers and listeners generate predictions of their own and their partner's utterances in order to successfully produce and comprehend speech (Dell & Chang, 2014; Federmeier, 2007; Pickering & Garrod, 2013; cf. Hickok, 2012). The engagement of coordination smoothers in conversations is in line with this view because any mechanism that makes utterances more predictable benefits linguistic prediction, thus making language processing easier.

Another related body of research concerns temporal synchronization in language. Many joint actions involve agents synchronizing their actions in time, and it has been proposed that such synchronization can serve as a coordination smoother (Vesper et al., 2010). Interestingly, in conversations, partners gradually converge on aspects of their language use including speech rate (Street, 1984; Schultz et al., 2016; Webb, 1969) and pausing patterns (Cappella & Planalp, 1981). Similarly, the study by Himberg et al. (2015), which we mentioned in the Introduction, found that speakers adapted to the speed with which their partner produced their last utterance. Based on a suggestion from an anonymous reviewer, we checked whether there was evidence for temporal adaptation in our joint pairs. Interestingly, we found that interacting participants progressively adapted the duration of their inter-turn intervals to one another (p = .010), but not the duration of their typing turns (p = .921; for further details, see the section Temporal Adaptation in Joint Pairs, in Supplement). These striking effects could be due to several mechanisms, but one possibility is that they stem from an inherent tendency of the cognitive system to entrain to the rhythms present in the environment (for a discussion, see Lelonkiewicz, Gambi, Weller, & Pfister, 2020). According to dynamic models of synchronization, the activity of one interacting agent becomes synchronized with the activity of another through a mechanism of perceptual coupling, similarly to two metronomes synchronizing via a mechanical connection (Dumas, de Guzman, Tognoli, & Kelso, 2014; Fairhurst, Janata, & Keller, 2012; Schmidt & O'Brien, 1997). Thus, it is possible that, due to perceptual coupling, speakers in a conversation may entrain some temporal aspects of their utterances (Wilson & Wilson, 2005). For example, Jungers and Hupp (2009) showed that the speech rate of comprehended utterances affects the speech rate of produced utterances, and Corps, Gambi, and Pickering (2020) recently showed that the speech rate of comprehended questions also affects the duration of the silent interval before an answer is produced.

Furthermore, the involvement of coordination smoothers in linguistic interactions is compatible with alignment accounts of dialogue (Pickering & Garrod, 2004). There is vast evidence that conversational partners align their language use on multiple levels, including tone of voice (Smith-Genthôs, Reich, Lakin, & de Calvo, 2015), pitch (Gregory & Webster, 1996), pronunciation (Pardo, 2006), accent (Giles, Coupland, & Coupland 1991), lexical choices (Branigan, Pickering, Pearson, McLean, & Brown, 2011) and the use of grammatical structures (Branigan, Pickering, & Cleland, 2000). According to the interactive alignment account, this tendency results from a bidirectional priming mechanism operating between language comprehension and production (Pickering & Garrod, 2004). For instance, comprehending a particular linguistic representation increases its activation in the language system, thus making the subsequent use of the same representation in production more likely (e.g., a person who just heard or read the word BUS is more likely to use the word BUS rather than COACH during their own speaking turn; Branigan, Pickering, Pearson, McLean, & Brown, 2011). Crucially, such priming-induced alignment percolates across different linguistic levels and affects all parties involved in a conversation, allowing the interacting speakers to align their representations for the conversations. In this view, adopting a common temporal pattern of language production could be a result of progressive interpersonal alignment which ultimately contributes to mutual understanding and so facilitates conversation.

While our findings are compatible with all above-mentioned accounts, it is important to note that more research is needed to clarify the functional significance of the reduction in the variability of inter-turn intervals. In joint motor tasks, variability reduction has been shown to lead to better coordination outcomes (i.e., more precise coordination of movements; Vesper et al., 2011; 2016). To explore whether there was a link between timing variability reduction and task success in our linguistic task, we asked whether reduced variability in inter-turn intervals would indicate that inter-turn intervals were shorter, as shorter gaps between turns are taken as an indication of coordination success in conversations (Levinson, 2016). There was indeed a clear positive association between variability and duration, but it should be noted that there was a similar association for typing times, even though we did not observe a significant reduction in temporal variability for this measure when comparing pairs to solo participants (ps < .001; for details, see The Relation Between Variability and Duration in Joint Pairs, in Supplement). Thus, this finding is suggestive but should be interpreted with caution.

In addition, for each pair in the joint condition we correlated timing variability with measures of definition quality, to explore whether those pairs who were more successful in reducing timing variability were also more successful at the task (we thank an anonymous reviewer for this suggestion). Interestingly, we found that variability in inter-turn intervals was associated with definition quality (p = .038), but variability in typing times was not (p = .769), suggesting that the timing between turns impacted on task performance more than the duration of the turns themselves (see The Relation Between Variability and Definition Quality in Joint Pairs, in Supplement). Surprisingly, however, pairs who exhibited *more* variability in inter-turn intervals were actually those who produced definitions of better quality. While the underlying cause of this relation is unclear, it may suggest that the adoption of a coordination smoother was not actually beneficial to task success in our task even though it made one aspect of the interaction - turn-taking - smoother.

We now turn to discussing the limitations of our methods. First, our measure of predictability was confined to the final word of the definition, meaning that it could not capture the predictability of the definition at earlier time points. Collecting predictability information for every contribution in every definition would have been impractical, and this is why we settled on only testing the predictability of the final contribution. However, we do acknowledge that this may reduce the sensitivity of this measure, because sentences are generally more predictable towards the end anyway, and so there may have been less scope for our manipulations to have an effect since we restricted our predictability analysis to this location.

Second, our measure of typing time includes both time spent typing and any pauses after the first keystroke, so it is difficult to tease apart how much of the variability relates to typing per se, and how much of it relates to utterance planning processes. However, we argue that, since there is no reason to assume that different experimental conditions affected the mechanical process of typing, any differences in typing time between conditions must reflect a change to the cognitive processes behind language production.

Finally, one may ask to what extent our findings can be translated onto different communicative situations. In particular, it is striking that participants in our task seemed to have engaged in utterance planning chiefly during their typing turn, rather than ahead of commencing production (i.e., during the gaps between turns). However, the amount of advance planning speakers engage in is known to be flexible (Konopka, 2012; Swets, Jacovina, & Gerrig, 2013; Van de Velde, Meyer, & Konopka, 2014), and it is possible that typed communication encourages a minimum amount of advance planning, because making corrections while already typing is arguably easier than while speaking, and because - in most typed environments, including the one used in our study - written contributions can be inspected and revised multiple times before they are made available to the listener. For this reason, our results may be particularly relevant for interactions involving dynamic written communication, for example on-line chats and conversing via instant messaging apps. Nevertheless, we do suggest that our task resembled spoken conversations in other critical ways, e.g., participants alternated comprehending and contributing to the exchange (Clark & Schaefer, 1989), they completed each other's sentences (Howes et al., 2011), and worked towards a common communicative goal (Clark, 1996).

In conclusion, our study contributes to the evidence suggesting that linguistic coordination is supported not only by listeners, but also speakers: We found that speakers engaged in a joint language production task initiated their turns after less variable delays than the individuals who performed the task on their own. This accords with our proposition that turn-taking in conversations may be facilitated by speakers making themselves more predictable. More broadly, our evidence is compatible with accounts suggesting that language processing in conversations involves processes of mutual coordination.

Acknowledgements

We would like to thank Gregory Mills for help in developing the version of DiET used in our task, and Lucia Castillo for contributing to participant testing. We would also like to thank Hugh Rabagliati and Martin Pickering for financially supporting part of this work from Leverhulme Trust Research Project Grant (RPG-2014-253).

References

Abdi, H. (2010). Coefficient of variation. In N. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 169-171). Thousand Oaks, CA: Sage.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264.

Armstrong, B. C., Tokowicz, N., & Plaut, D. C (2012). eDom: Norming software and relative meaning frequencies for 544 English homonyms. *Behavior Research Methods, 44*, 1015–1027.

Barthel, M., Meyer, A. S., & Levinson, S. C. (2017). Next speakers plan their turn early and speak after turn-final “go-signals”. *Frontiers in Psychology, 8*, 393.

Barthel, M., & Sauppe, S. (2019). Speech Planning at Turn Transitions in Dialog Is Associated with Increased Processing Load. *Cognitive Science, 43*(7), e12768.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2019). lme4: Linear mixed-effects models using Eigen and S4. http://CRAN.R-project.org/package=lme4

Boiteau, T. W., Malone, P. S., Peters, S. A., & Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *Journal of Experimental Psychology: General, 143*(1), 295-311.

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports, 5*, 12881.

Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics, 52*, 46-57.

Branigan, H. P., Catchpole, C. M., & Pickering, M. J. (2011). What makes dialogues easy to understand? *Language and Cognitive Processes, 26*(10), 1667-1686.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*(2), 13-25.

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition, 121*(1), 41-57.

Cappella, J. N., & Planalp, S. (1981). Talk and silence sequences in informal conversations III: Interspeaker influence. *Human Communication Research, 7*(2), 117-132.

Clark, H. H. (1996). *Using language*. Cambridge university press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*(2), 259-294.

Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. *Cognition, 175*, 77-95.

Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes, 55*(2), 230-240.

Corps, R. E., Gambi, C., & Pickering, M. J. (2020). How do listeners time response articulation when answering questions? The role of speech rate. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(4), 781-802.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391-407.

Dumas, G., de Guzman, G. C., Tognoli, E., & Kelso, J. S. (2014). The human dynamic clamp as a paradigm for social interaction. *Proceedings of the National Academy of Sciences, 111*(35), E3726-E3734.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology, 23*(2), 283.

Fairhurst, M. T., Janata, P., & Keller, P. E. (2012). Being and feeling in sync with an adaptive virtual partner: Brain mechanisms underlying dynamic cooperativity. *Cerebral Cortex, 23*(11), 2592-2600.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469-495.

Fjaellingsdal, T. G., Schwenke, D., Scherbaum, S., Kuhlen, A. K., Bögels, S., Meekes, J., & Bleichner, M. G. (2020). Expectancy effects in the EEG during joint and spontaneous word-by-word sentence production in German. *Scientific Reports*, *10*(1), 1-12.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: quantifying the benefits of linguistic coordination. *Psychological science*, *23*(8), 931-939.

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science, 1*(2), 292-304.

Giles, H., Coupland, J., & Coupland, N. (1991). *Contexts of accommodation: Developments in applied sociolinguistics*. New York: Cambridge University Press.

Gravano, A., & Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech & Language, 25*(3), 601-634.

Gregory, S. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status. *Journal of Personality and Social Psychology, 70*(6), 1231-1240.

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun – An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods, 47*, 930-944.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics, 38*(4), 555-568.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13*, 135-145.

Himberg, T., Hirvenkari, L., Mandel, A., & Hari, R. (2015). Word-by-word entrainment of speech rhythm during joint story building. *Frontiers in Psychology*, *6*, 797.

Howes, C., Purver, M., Healey, P. G., Mills, G., & Gregoromichelaki, E. (2011). On incrementality in dialogue: Evidence from compound contributions. *Dialogue & Discourse, 2*(1), 279-311.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1-2), 101-144.

Jungers, M. K. & Hupp, J. M. (2009). Speech priming: Evidence for rate persistence in unscripted speech. *Language and Cognitive Processes, 24*, 611-624.

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: theory and data. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 59-101). Academic Press.

Konopka, A. E. (2012). Planning ahead: How recent experiences with structures and words changes the scope of linguistic planning. *Journal of Memory and Language, 66*, 143–162.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190-207). Oxford Scholarship Online.

Lelonkiewicz, J. R., Gambi, C., Weller, L., & Pfister, R. (2020). Action–effect anticipation and temporal adaptation in social interactions. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(4), 335-349.

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in Cognitive Sciences, 20*(1), 6-14.

Lindsay, L., Gambi, C., & Rabagliati, H. (2019). Preschoolers Optimize the Timing of Their Conversational Turns Through Flexible Coordination of Language Comprehension and Production. *Psychological Science, 30*(4), 504-515.

Magyari, L., Bastiaansen, M. C., de Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience, 26*(11), 2530-2539.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314-324.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture–word interference experiments. *Journal of Memory and Language, 35*(4), 477-496.

Meyer, M., Bekkering, H., Haartsen, R., Stapel, J. C., & Hunnius, S. (2015). The role of action prediction and inhibitory control for joint action coordination in toddlers. *Journal of Experimental Child Psychology, 139*, 203-220.

Mills, G. J. (2014). Dialogue in joint activity: Complementarity, convergence and conventionalization. *New Ideas in Psychology, 32*, 158-173.

Mills, G. J., & Healey, P. G. T. (submitted). A dialogue experimentation toolkit. Retrieved from http://cogsci.eecs.qmul.ac.uk/diet/

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America, 119*(4), 2382-2393.

Pezzulo, G., Donnarumma, F., & Dindo, H. (2013). Human sensorimotor communication: A theory of signaling in online social interactions. *PLoS ONE, 8*(11)*, p. e79876.*

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002-1044.

Pickering, M. J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, *27*(2), 212-225.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain sciences*, *36*(4), 329-347.

Richardson, D., Dale, R., & Shockley, K. (2008). Synchrony and swing in conversation: Coordination, temporal dynamics, and communication. In I. Wachsmuth, M. Lenzen, and G. Knoblich (Eds.), *Embodied Communication in Humans and Machines* (pp. 75-94). Oxford Scholarship Online.

Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology, 2*(307), 1–14.

Schultz, B. G., O'Brien, I., Phillips, N., & McFarland, D. H. (2016). Speech rates converge in scripted turn-taking conversations. *Applied Psycholinguistics, 37*, 1201-1220.

Sjerps, M. J., & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition*, *136*, 304-324.

Skantze, G., Hjalmarsson, A., & Oertel, C. (2014). Turn taking, feedback and joint attention in situated human-robot interaction. *Speech Communication, 65*, 50-66.

Schmidt, R. C., & O'Brien, B. (1997). Evaluating the dynamics of unintended interpersonal coordination. *Ecological Psychology, 9*(3), 189-206.

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition, 73*(3), 205-246.

Smith, M., & Wheeldon, L. (2004). Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(3), 675–686.

Smith-Genthôs, K. R., Reich, D. A., Lakin, J. L., & de Calvo, M. P. C. (2015). The tongue-tied chameleon: The role of nonconscious mimicry in the behavioral confirmation process. *Journal of Experimental Social Psychology, 56*, 179-182.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences, 106*(26), 10587-10592.

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research, 11*, 139-169.

Swets, B., Jacovina, M. E., & Gerrig, R. J. (2013). Effects of conversational pressures on speech planning. *Discourse Processes, 50*, 23–51.

Taylor, W. L. (1953). “Cloze procedure”: A new tool for measuring readability. *Journalism Quarterly*, *30*(4), 415-433.

Tolins, J., Zeamer, C., & Fox Tree, J. E. (2018). Overhearing dialogues and monologues: How does entrainment lead to more comprehensible referring expressions? *Discourse Processes, 55*(7), 545-565.

Van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing “easy” and “hard” events with preferred and dispreferred syntactic structures. *Journal of Memory and Language, 71*, 124–144.

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, *23*(8-9), 998-1003.

Vesper, C., Schmitz, L., Sebanz, N., & Knoblich, G. (2013). Joint action coordination through strategic reduction of variability. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1522–1527). Austin, TX: Cognitive Science Society.

Vesper, C., Schmitz, L., Safra, L., Sebanz, N., & Knoblich, G. (2016). The role of shared visual information for joint action coordination. *Cognition, 153*, 118-123.

Vesper, C., van der Wel, R. P., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable: reduced temporal variability facilitates joint action coordination. *Experimental Brain Research, 211*(3-4), 517-530.

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(2), 423-440.

Webb, J. T. (1969). Subject speech rates as a function of interviewer behaviour. *Language and Speech, 12*(1), 54–67.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, *131*(3), 460-473.

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review, 12*(6), 957-968.

Supplementary Materials

Contents

Additional Analyses

Following the suggestions of anonymous reviewers, we carried out the analyses outlined below. These exploratory analyses offer further insights into our data.

Typing Time and Inter-Turn Interval Duration

Our primary analysis focused on the temporal variability with which participants produced definitions: We found that joint pairs showed less variability in inter-turn intervals as compared to solo participants, and that pairs' variability in typing times was greater for definitions produced for ambiguous than unambiguous words.

But in addition, one could ask whether similar effects of Partner Condition and Ambiguity would emerge with regards to the *duration* of typing times and inter-turn intervals. To test this, we re-ran the LME used for our primary analysis, but this time regressing the duration of inter-turn intervals (i.e., average lag between turns computed over each definition produced by the solo participant/pair), as well as ran an analogous LME regressing the duration of typing times (i.e., average time spent typing a word computed over each definition produced by the solo participant/pair). The models followed the syntax: Interval/Typing Time Duration ~ 1 + Ambiguity * Partner Condition + Length + (1 + Ambiguity || Author) + (1 + Partner Condition | Mystery Word). The numerical trends are illustrated in Figure S1 and model summaries are reported in Table S1.

Ambiguity as a manipulation of task difficulty. The analysis found no significant effects of Ambiguity on either inter-turn interval or typing time duration, suggesting that participants spent comparable amounts of time defining both ambiguous and unambiguous words. This prompts the question whether our ambiguity manipulation was an effective way to manipulate task difficulty - after all, greater difficulty should in principle cause participants to slow down. With the caveat that null effects should be interpreted with caution, a possible explanation for the lack of a reliable Ambiguity effect on duration is that participants responded to this manipulation by slowing down overall, that is, for both ambiguous and unambiguous words (e.g., after noticing that the partner is defining a different meaning of the word participants could become generally more cautious, and hence slower). Alternatively, it is possible that ambiguity led to increased task difficulty at particular points in a definition (e.g., in the beginning) rather than to an overall slow-down across the entire definition. This would result in increased durations for a few intervals/turns, which would not have affected the average duration of the definition as much, but would have been reflected in increased variability as measured by the coefficient of variation (see manuscript main text).

Importantly, all our other measures were affected by the ambiguity manipulation in the expected direction (see main text): We found that the definitions produced for ambiguous words were characterized by lower quality and lower predictability, as compared to the definitions produced for unambiguous words, and that joint pairs showed more variability in typing times when defining ambiguous than unambiguous words. Thus, our findings were overall consistent with the hypothesis that defining ambiguous mystery words was indeed associated with increased difficulty.

Speed-accuracy trade-offs. The duration analysis also revealed that solo participants were significantly faster than joint pairs (for both inter-turn intervals and typing times). This result is informative with regards to another potential interpretational issue: Recall that we found that definitions produced in the joint condition were of poorer quality than those produced by solo participants (see Figure 5 in the main text). One could ask whether this difference could stem from a trade-off between speed and accuracy - participants in the joint condition may have been socially pressured to be faster than participants in the solo condition, which in turn could have led them to produce worse definitions (we thank an anonymous reviewer for suggesting this point). Importantly, the fact that joint participants were actually slower than their solo counterparts suggests that the poorer quality of joint definitions is unlikely to be due to a speed-accuracy trade-off.

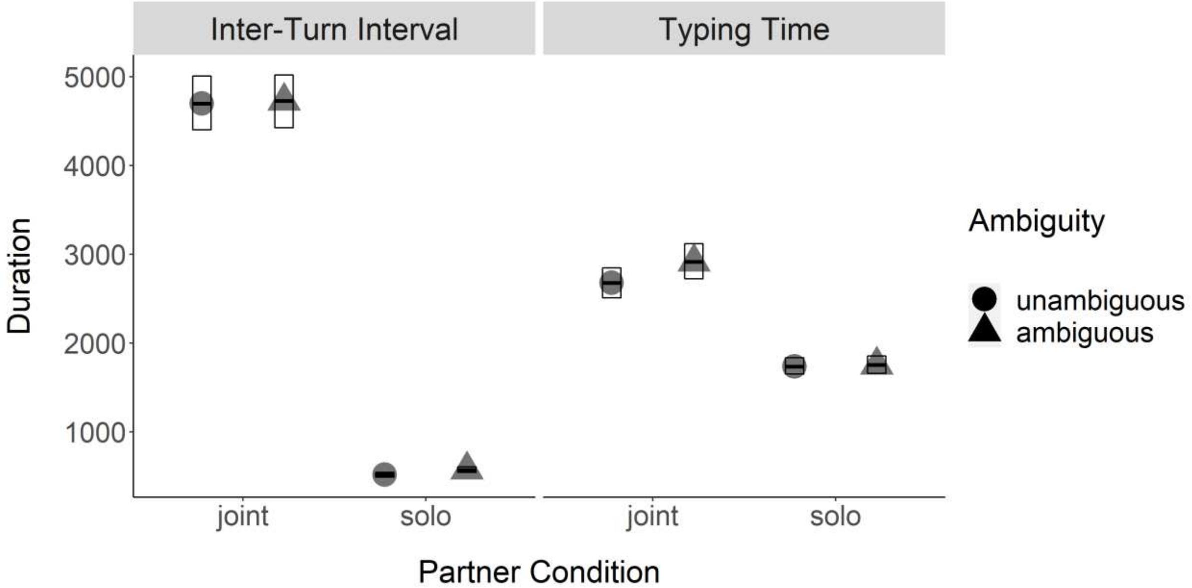

Figure S1. Mean duration of Inter-Turn Intervals and Typing Times, presented by Partner Condition (joint vs. solo) and Ambiguity (unambiguous vs. ambiguous). The error bars represent 95% confidence intervals.

Table S1. Summary of linear mixed-effect models for the durations of Inter-Turn Intervals and Typing Times.

Inter-Turn Interval Duration				
Fixed effect	*B*	*SE*	*t*	*p*
Length	10.21	10.49	0.97	0.331
Partner Condition	4202.89	274.12	15.33	< 0.001
Ambiguity	41.76	168.65	0.25	0.806
Partner Condition:Ambiguity	-29.15	381.79	-0.08	0.940
Random effect				Variance
Author				399446
Mystery Word				216721
Author:Ambiguity				29743
Mystery Word:Partner Condition				1177219
Typing Time Duration				

Fixed effect	*B*	*SE*	*t*	*p*
Length	-26.18	8.26	-3.17	0.001
Partner Condition	957.23	198.09	4.83	< 0.001
Ambiguity	96.23	121.58	0.79	0.434
Partner Condition:Ambiguity	233.45	169.14	1.38	0.176
Random effect				Variance
Author				329159
Mystery Word				110425
Author:Ambiguity				< 0.001
Mystery Word:Partner Condition				131131

The Relation Between Variability and Duration in Joint Pairs

Vesper et al. (2011; 2016) reported a positive relationship between variability (i.e., standard deviation; SD) and reaction times in motor joint tasks. We investigated if the same relation can be observed in our linguistic task in the joint condition. To do so, we ran a LME on inter-turn interval duration with standard deviation in these intervals as a predictor, and another LME on typing times duration with standard deviation in typing times as a predictor. To help convergence, the predictors were scaled and centred. The models followed the syntax: Interval/Typing Time Duration ~ 1 + Interval SD/Typing Time SD + (1 | Author) + (1 | Mystery Word). The model summaries are reported in Table S2.
Interestingly, we found a clear positive association between variability and typing time/inter-turn interval durations. This lends some support to Vesper et al. (2011; 2016) who proposed that interaction partners may adopt quicker response times in order to reduce temporal variability. However, recall that pairs reduced their variability with regards to inter-turn intervals, but not typing times, compared to solo participants. Thus, this positive relation may, at least in part, be also explained by a floor effect - the shorter the durations, the more constrained variability. Further research is needed to understand the underlying mechanisms.

Table S2. The relationship between variability and duration in joint pairs: Summary of linear mixed-effect models for mean duration of inter-turn intervals and typing times with variability (SD) as a predictor.

Inter-Turn Interval Duration				
Fixed effect	*B*	*SE*	*t*	*p*
Interval SD	2355.63	51.98	45.32	< 0.001
Random effect				Variance

Author				108412
Mystery Word				206066
Typing Time Duration				
Fixed effect	*B*	*SE*	*t*	*p*
Typing Time SD	1661.70	28.47	58.36	< 0.001
Random effect				Variance
Author				81439
Mystery Word				33623

The Relation Between Variability and Definition Quality in Joint Pairs

Our primary analysis revealed that joint pairs engaged in temporal variability reduction. But why? It is of course possible that this mechanism might be routinely engaged in joint actions, as a way of facilitating coordination. However, another possibility is that reducing one's temporal variability specifically served the goal of completing the task we set to participants (i.e., to jointly produce a good quality definition). To investigate this possibility, we tested if better coordinated (i.e., less variable) pairs produced definitions of higher quality. For each joint pair, we calculated the average quality of the definitions produced by that pair (we did that for either of our quality measures, i.e., explicit quality ratings and proportion of correct guesses of the mystery word), the average length of the produced definitions, as well as the average temporal variability of that pair (i.e., inter-turn intervals COV, typing times COV). We then ran linear regressions on either quality measure, with both variability measures as predictor (scaled and centred), and controlling for definition length (scaled). The models followed the syntax: Quality Rating/ACC Mystery Word Guess ~ 1 + Interval COV + Typing Time COV + Length. The summary of fixed effects from the models can be found in Table S3.
Interestingly, we found that variability in inter-turn intervals was correlated with definition quality, but variability in typing times was not, suggesting that the timing between turns impacted on task performance more than the duration of the turns themselves. Surprisingly, however, pairs who exhibited *more* variability in inter-turn intervals were actually those who produced definitions of better quality. While the underlying cause of this relation is unclear, it may suggest that the adoption of a coordination smoother was not actually beneficial to task success in our task even though it made one aspect of the interaction - turn-taking - smoother.

Table S3. The relation between variability and definition quality in joint pairs: Summary of linear regression models investigating the effect of variability measures on definition quality, controlling for definition length.

Explicit Quality Rating				
Fixed effect	*B*	*SE*	*t*	*p*
Interval COV	0.054	0.05	0.98	0.344
Typing Time COV	-0.054	0.06	-0.90	0.384
Length	-0.003	0.06	-0.05	0.959
ACC Mystery Word Guess				
Fixed effect	*B*	*SE*	*t*	*p*
Interval COV	0.037	0.02	2.29	0.038
Typing Time COV	-0.005	0.02	-0.30	0.769
Length	-0.023	0.02	-1.27	0.225

Temporal Adaptation in Joint Pairs

The evidence from our primary analysis supported the hypothesis that linguistic coordination benefits from a mechanism of variability reduction. However, another way of making the coordination smoother is for partners to dynamically adapt to each other's response speed (e.g., Knoblich et al., 2011; Marsh, Richardson, Schmidt, 2009; Vesper et al., 2010). We investigated if our joint pairs engaged in such adaptation by testing if participant's typing time/inter-turn interval on the current turn was predicted by their partner's typing time/inter-turn interval on preceding turn, and whether this relationship was modulated by trial (i.e., the order of the turn within a definition - whether it was the first, second, third, …, n-th word produced as part of that definition). To improve convergence, typing times/intervals were scaled and centred, trial was scaled. We specified two separate LME models of the following syntax: Typing Time/Interval ~ 1 + Preceding Typing Time/Preceding Interval * Trial + (1 | Author) + (1 | Mystery Word). The results are reported in Table S4.

In line with previous studies (e.g., Lelonkiewicz & Gambi, 2017), the analysis revealed a positive relationship between the current *inter-turn interval* (i.e., time between turn onset and action initiation of the current speaker) and partner's interval on the preceding turn, suggesting that joint pairs entrained their interval durations. Further, such entrainment grew stronger as participants progressed through a definition, as indicated by a statistically reliable interaction between Preceding Interval and Trial. These findings lend some support to the proposition that linguistic coordination is bolstered also by temporal adaptation (Wilson & Wilson, 2005).

But curiously, we found no evidence for temporal adaptation defined as similarity between partners' *typing times* - neither the effect of Preceding Typing Time, nor the interaction between Preceding Typing Time and Trial were statistically significant. Together with the results from our variability analysis (i.e., joint pairs reduced variability in inter-turn intervals, but failed to do so for typing times), this again suggests that participants in our task focused on inter-turn intervals in their use of coordination smoothers (we discuss possible reasons for that in the main text).

Finally, the temporal adaptation analysis also revealed that participants became quicker as they were nearing the completion of a definition (i.e., a significant negative correlation with Trial with regards to both inter-turn intervals and typing times), possibly because later trials were characterized by more coordination (closer temporal adaptation) than the first few trials in a definition.

Table S4. Temporal adaptation in joint pairs: Summary of linear mixed-effect models investigating temporal adaptation in typing times and inter-turn intervals.

Inter-Turn Interval				
Fixed effect	*B*	*SE*	*t*	*p*
Preceding Interval	0.103	0.03	3.94	< 0.001
Trial	-0.221	0.03	-8.00	< 0.001
Preceding Interval:Trial	0.069	0.02	2.57	0.010
Random effect				Variance
Author				0.02804
Mystery Word				0.89555

Typing Time				
Fixed effect	*B*	*SE*	*t*	*p*
Preceding Typing Time	0.002	0.05	0.09	0.927
Trial	-0.214	0.03	-7.61	< 0.001
Preceding Typing Time:Trial	-0.002	0.02	-0.10	0.921
Random effect				Variance
Author				0.025023
Mystery Word				0.009249

Descriptive Statistics: Content Predictability and Quality

As reported in the main text, we tested whether the definitions produced by joint pairs differed from those produced by solo participants in terms of content predictability and quality, and whether the definitions produced for ambiguous mystery words differed along these dimensions from those produced for unambiguous words. In the context of these analyses, one could ask whether predictability and quality measures were characterized by sufficient variability to allow for a reliable detection of statistical effects.

As illustrated in the plots below (Figures S2 and S3), there was a considerable variability in both predictability (accuracy of final word guesses: range = 0-55%, mean = 28%; LSA scores: range = 0.02-0.60, M = 0.35; prior to calculating these statistics, we excluded one high outlier observation) and quality (explicit quality ratings: range = 2.15-4.70, M = 3.45; accuracy of mystery word guesses: range = 30-95%, M = 65%; prior to calculating the statistics for guesses accuracy, we excluded one low outlier). Thus, it is unlikely that the analyses of predictability and quality were biased by insufficient variability in these variables.

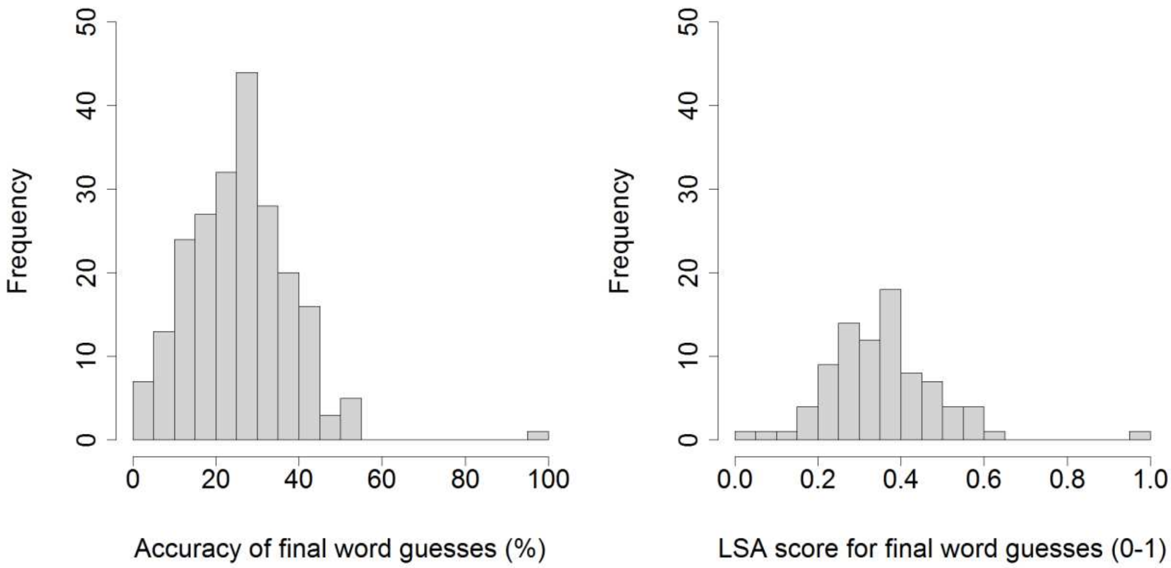

Figure S2. Content predictability: Histograms for accuracy of final word guesses and Latent Semantic Analysis (LSA) scores of these guesses, summarised by Partner Condition and Ambiguity for each participant.

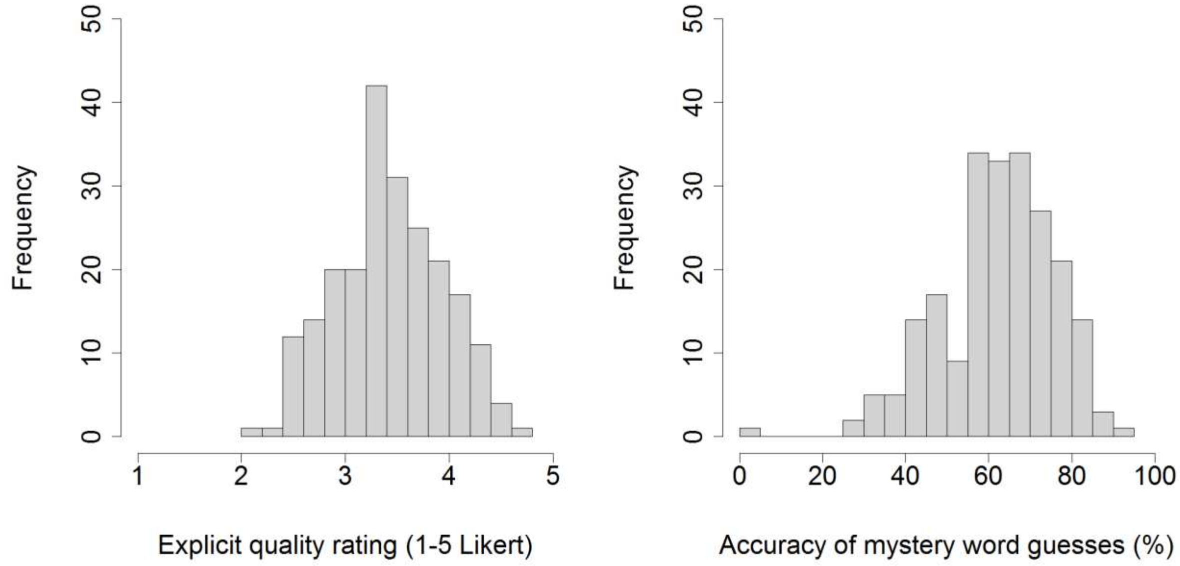

Figure S3. Content quality: Histograms for explicit quality ratings and accuracy of mystery word guesses, summarised by Partner Condition and Ambiguity for each participant.

Supplementary References

Lelonkiewicz, J. R., & Gambi, C. (2017). Spontaneous adaptation explains why people act faster when being imitated. *Psychonomic Bulletin & Review*, *24*(3), 842-848.

Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science*, *1*(2), 320-339.