Joint language production and the representation of other speakers' utterances.

Chiara Gambi

Cardiff University

Martin J. Pickering

University of Edinburgh

Address for correspondence:

Chiara Gambi

School of Psychology

70, Park Place

Cardiff University

CF10 3AT Cardiff, U.K.

Email: GambiC@cardiff.ac.uk

Joint language production and the representation of other speakers' utterances.

Joint language production is the study of the mechanisms involved in producing language jointly with another real or assumed speaker. Using relevant tasks, researchers have asked how (if at all) speakers represent one another's utterances (which we term *co-representation*) and specifically examined how such representations affect language production. They have compared the process of producing language jointly to the process of producing language individually, and have done so to address a question that is important for the study of both comprehension and production: How do language production processes relate to the representation of others' utterances? If production processes contribute to co-representation, then we would expect to find that co-representation affects actual language production.

We use the term *joint language production* to refer both to cases where two people speak at the same time and to cases where people take turns speaking (e.g., *A* names a picture, then *B* names a picture). Furthermore, we include both situations where speakers are simply aware of each other's tasks (e.g., Gambi et al., 2015a; Kuhlen & Abdel Rahman, 2017) and situations where speakers intend to coordinate with each other, such as in choric production (Cummins, 2003, 2009; Jasmin et al., 2016), or when they are instructed to jointly construct a meaningful sentence (e.g., Lelonkiewicz & Gambi, 2020) or to minimize the silent pause between their utterances (e.g., Hoedemaker & Meyer, 2018). Importantly, we use it to refer not only to situations where two people are actually producing, but also to instances when one participant produces while believing that another person is also producing (whether the participant receives any feedback about their partner's production or not, or even whether there is a real partner or not). Finally, speaking individually (the other side of the comparison) can refer either to one speaker performing just half of the joint production task (e.g., *A* names a picture, but then nobody speaks) or to one speaker performing the whole of the task (e.g., *A* names a picture, and then names another picture). Throughout the chapter, we specify exactly what is meant by "jointly" and "individually" with reference to particular studies, but we first explain the theoretical importance of this comparison and thus the unique contribution made by this literature.

Comparing joint to individual production is important because it helps answering the following question: To what extent do a speaker's language production processes contribute to the representation of another speaker's utterances? Such representation is typically considered to be part of comprehension, but traditional studies of comprehension do not consider the extent to which others' utterances are represented similarly to people's own utterances (as an example, see the chapters in Crocker et al., 2000). The assumption of considerable overlap between representations and processes used in production and comprehension (Gambi & Pickering, 2017) is at the heart of some theories of dialogue (Pickering & Garrod, 2013; Pickering & Garrod, 2021), monitoring (see Gauvin & Hartsuiker, 2020 for a recent review), and prediction (Pickering & Gambi, 2018). However, until recently, very little was known about how the processes that underlie language production in an individual are adapted to incorporate representations of others' utterances in a joint setting. Furthermore, the hypothesis that speakers use the production system to represent the utterances of other speakers as if they were their own (i.e., via simulation; Dell & Chang, 2014; Pickering & Garrod, 2013) has remained controversial (Hickok, 2013).

Note that our question is more specific than the broad question of how others' utterances affect one's own language production. The latter is of course central to the study of dialogue: In referential games (e.g., Brennan & Clark, 1996; Garrod & Anderson, 1987), in work about the effect of feedback and backchannels (e.g., Bavelas et al., 2000; Tolins et al., 2018), in studies of syntactic priming (e.g., Branigan et al., 2000) and priming of language switching across interlocutors (Kootstra et al., 2010), and more recently in studies of turn-taking (e.g., Bögels et al., 2015; Corps et al., 2018), researchers are interested in how an interlocutor's production (lexical and structural choices, language choices, the time course of sentence preparation) is affected by another speaker's utterances.

Most studies of dialogue are not designed to determine whether production processes are involved in the representation of others' utterances, and thus do not include direct comparisons between effects of other's utterances and one's own previously produced utterances on language production. A few studies of syntactic priming (i.e., the tendency of speakers to re-use recently comprehended or produced structural representations; see Pickering & Ferreira, 2008; Slevc, this volume) have included a comparison between production-to-production priming (i.e., priming within an individual speaker)

and comprehension-to-production priming (i.e., priming between speakers) (Bock et al., 2007; Jacobs et al., 2019; Segaert et al., 2013). But they tested isolated speakers (i.e., the other speaker on comprehension-to-production trials was just implied) and it is not clear to what extent the production task used on prime trials reflected only production (and not comprehension) processes (Jacobs et al., 2019).

We are not aware of any experimental study of priming in dialogue that has directly compared within-speaker and between-speaker priming. In contrast, some corpus studies have included this comparison (e.g., Gries, 2005; Reitter et al., 2006). While they found stronger within- than between-speaker priming, it is difficult to be certain about the relationship between comprehension and production mechanisms because the findings could be confounded by other factors such as discourse structure (see Jacobs et al., 2019 for discussion). Finally, Schoot et al. (2019) found larger priming in the presence versus absence of an interlocutor (though see Ivanova et al., 2020, for contrasting findings), but the source of this effect is unclear: The presence of an interlocutor may have encouraged production-based simulation of the primes during comprehension, but the study did not include a comparison to production-to-production priming.

In this chapter we review studies that include direct comparisons between individual and joint production (sometimes, we make brief reference to studies that included only a joint production condition). Taken together, the studies show that language production mechanisms are affected by representations of others' utterances: Speakers represent whether their co-speakers' are engaging in language production (i.e., speaking or preparing to speak) and these representations influence the way they produce their own utterances, often in a way that parallels within-speaker effects. However, the extent to which this co-representation alters the dynamics of language production compared to speaking individually appears limited, such that speakers are typically affected by others' utterances in a less specific way or to a lesser extent than by their own utterances. These findings show that representation of others' utterances and production of one's own utterances are based on partly overlapping mechanisms, but also that representations of others' utterances tend to be less detailed.

Below, we organise our review broadly by paradigm. We identify three individual production paradigms for which researchers have developed joint versions: (1) picture naming (including the word-

replacement task used by Gambi et al., 2015a – see below); (2) picture-word interference (PWI) and Stroop tasks; (3) language switching tasks. We then briefly review two special cases (4): choric production and joint sentence building. In our conclusion, we consider open questions and the relevance of this research for the study of dialogue.

1. Joint picture naming

Most of the studies that have investigated joint language production have used variations of the picture naming paradigm, in which speakers are asked to name pictures displayed on a screen. Some details, for instance the number of pictures displayed simultaneously, or whether participants sat next to each other or in separate rooms, varied between studies. But in all of them speakers either named pictures on their own or shared the task with another speaker, who might be a (naïve) participant or a confederate.

In one of the earliest studies, Gambi, Cop, and Pickering (2015a) had two participants sit next to each other in front of the same screen and take turns naming pictures. On critical trials, a picture changed into a different picture, and participants were instructed to stop speaking as quickly as possible. One group was told that the person who stopped would name the new picture (individual task group), another group was told that the other person would name the new picture (joint task group), and a third group had to ignore the new picture (control group). Participants in the individual task group were less likely to stop mid-word than participants in the control group (replicating Hartsuiker et al. (2008) and Tydgat et al. (2011) in the presence of a partner), thus suggesting that planning the new picture name interfered with the process of stopping speech. Importantly, participants in the joint task group were also less likely to stop mid-word than participants in the individual task group, but not to the same extent as participants in the control group. Thus representing that another person will speak delays the process of stopping speech similarly to representing that one will speak, albeit to a lesser extent.

Two studies manipulated speakers' beliefs about the task their partner was performing in a different room while they prepared to speak (Gambi et al., 2015b; 2022). In Gambi et al. (2015b), speakers named either single pictures or superimposed pairs of pictures and, across four experiments, they were told their partner would produce an utterance which was the same or different from their own

utterance, stay silent, or respond yes or no to a semantic categorization question (i.e., did the pictures belong to the same semantic category?). Naming latencies were faster when participants believed their partner was not speaking, or was speaking but not engaging in lexical retrieval (categorization condition), than when they believed their partner was naming pictures. Moreover, Gambi et al. (2022) replicated this finding in a task where speakers produced full active or passive descriptions of transitive events: Across three experiments, description latencies were longer when speakers believed their partner was producing or about to produce a sentence, compared to when they believed their partner remained silent.

Thus, speakers represented whether their partner was engaging in language production – even though doing so brought no obvious benefit to their performance, and in fact slowed down concurrent production. The fact that such representations affected concurrent language production suggests that co-representation of others' utterances makes use of production mechanisms. However, both studies also suggest that co-representation is only partial. Onset latencies were unaffected by whether speakers believed their partner was naming the same or a different picture (or indeed whether they were naming the pictures in the same or a different order; Gambi et al., 2015b) and they were similarly unaffected whether they believed the partner was describing the same event with a syntactic construction of the same or opposite voice (Gambi et al., 2022). There was some evidence that believing the partner was producing a different utterance than one's own increased interference (sometimes leading to increased error rates; Gambi et al., 2015b; or longer descriptions; Gambi et al., 2022), but it was not consistent across experiments, suggesting that co-representation generally lacked detail.

One important question is whether speakers may co-represent others' utterances in greater detail under different conditions than the one tested by Gambi and colleagues (2015b; 2022). First, in those studies, speakers performed a joint task only in a minimal sense: They sat in different rooms and had no requirement to coordinate their utterances with those of their partner. Speakers might be more likely to represent the content of their partner's utterances when their partner's presence is made more salient, or when they are given explicit instructions to coordinate. Second, speakers were co-representing others' utterances while preparing to speak themselves; but since language production is cognitively demanding (Roelofs & Piai, 2011), the resources available for concurrent co-representation

may have been limited. It is thus possible that speakers may represent their partners' utterances in greater detail when they do not need to speak at the same time as their partner, but take turns with them. Below, we review studies that help address these questions.

Brehm, Taschenberger, and Meyer (2019) manipulated speakers' beliefs about their partner's task (naming the same picture, a different picture, or categorizing it as living/non-living), similarly to Gambi et al. (2015b). However, they had participants sit side-by-side. Thus, although speakers wore noise-cancelling headphones, the co-present partner was arguably more salient. Despite this, there were again only inconsistent effects of the content of the partner's utterance. Across two experiments, participants took *longer* to name the pictures when they knew their partner was performing a different task (categorization) than the same task (naming), and this effect was present in the joint task (partner present) but not in an individual control task (when speakers had the same task instructions and stimuli, but were told their partner could not attend the session). Although one experiment found longer latencies when the partner was naming a different than the same picture, this effect was present even in the individual condition, suggesting it was not related to co-representation, and furthermore it was not replicated in the second experiment.

In sum, Brehm et al. (2019) found evidence for co-representation of the partner's task (though the effect was in the opposite direction to Gambi et al., 2015b), but not that they formed detailed representations of the partner's utterances, even when the partner's presence was salient. However, partners' pictures were displayed on the opposite side of the screen (cf. Gambi et al., 2015b, who superimposed the pictures), making it harder for speakers to inspect them (as confirmed by eye-tracking data). Thus, although speakers knew where the partner's pictures would be displayed and although they were told whether the partner's pictures were the same or different from theirs, they may have disregarded this information as task-irrelevant.

More compelling evidence against detailed co-representation, even when it is task relevant, comes from Hoedemaker and Meyer (2018). In two experiments, they had two speakers sit side by side, while one of the speakers was eye-tracked. Three pictures were displayed on each trial, and instructions either required only one participant to name one or more pictures (individual naming) or both participants to take turns (joint naming). When more than one picture was named, speakers were

instructed to minimize the silent pause between words. In joint naming, this meant coordinating the production of the utterances with their partner. Speakers could do this well, achieving gaps of comparable lengths to those found in natural conversations (below 300ms on average; Stivers et al., 2009). But although speakers were more likely to look at pictures their partner would later name (in the joint condition) than pictures nobody would name (in the individual condition), those looks were much shorter and closer to speech onset than looks to the same pictures when the speakers themselves would later name them, implying that co-representation stopped well short of planning the partner's utterances. Crucially, this was the case even when the task required speakers to coordinate their utterances, suggesting co-representation lacks detail even when representing the content of others' utterances could facilitate smooth turn-taking.

However, all the studies mentioned so far investigated whether speakers co-represent others' utterances while speaking or preparing to speak themselves. As mentioned above, concurrent production may limit the cognitive resources that speakers have available for co-representation. Thus, we should also consider studies that have tested whether speakers co-represent others' utterances in detail when they remain silent. In order to assess whether co-representation makes use of language production processes, these studies adopted two different approaches. One approach was to look for neural signatures of production processes on trials where the participant did not speak, but their partner did (Baus et al., 2014). The other was to examine onset latencies on subsequent trials on which the participant spoke to look for evidence of co-representation on previous silent trials (Hoedemaker et al., 2017; Kuhlen & Abdel Rahman, 2017, 2021).

Baus et al. (2014) had participants take turns producing high- and low-frequency picture names with a confederate, while their EEG was recorded (there was also an individual condition, tested in a separate block, with no confederate present). In line with previous literature (e.g., Strijkers et al., 2010), the amplitude of the P200 component was larger when participants prepared to produce low- than high-frequency words – a standard frequency effect. Crucially, there was also a frequency effect when the participant remained silent and the confederate prepared to speak, but not when the confederate stayed silent as well (or indeed in the individual condition, when there was no confederate), suggesting speakers performed lexical access for pictures that their partner was about to name. Interestingly, the

frequency effect on silent trials was delayed compared to speaking trials (frequency affected the P300, rather than the P200 component), which suggests that the processes underlying co-representation are similar to those underlying production, but slower.

Kuhlen and Abdel-Rahman (2017) used a joint version of the cumulative semantic inhibition paradigm. In the (standard) individual version of this paradigm, participants name pictures that belong to a limited number of semantic categories, and naming latencies become slower with each naming instance within a category (Brown, 1981). This effect is thought to result from changes in the strength of connections between conceptual and lexical representations (or between features and concepts), which are caused by previous retrieval episodes (Belke, 2013; Howard et al., 2006; Oppenheim,et al., 2010). In the joint version, a coloured frame around pictures indicated whether they should be named by the speaker, named by the partner, or named by neither; the speaker named half of the pictures in each semantic category, but crucially for some categories the partner named the other half of the pictures (joint naming categories), while for other categories half of the pictures were presented but not named by either the speaker or the partner (individual naming categories).

Latencies increased more steeply with successive naming instances for joint than individual naming categories. Crucially, this was the case not only when speakers could hear their partners (as also shown by Hoedemaker et al., 2017), but also when speakers could not hear their co-present partner (because they wore noise-cancelling headphones) or when the partner was seated in a different room (though see Kuhlen & Abdel Rahman, 2021, who did not replicate this last finding). In sum, there was evidence for between-speaker cumulative semantic inhibition, suggesting that speakers co-represented their partner's words using their production system, and that these co-representations led to changes in the strength of connections in the mental lexicon that are qualitatively similar to those that occur when speakers produce the words themselves. We do not know how this between-speaker effect compares to the within-speaker effect quantitatively (as the study did not include a condition where the speaker named all pictures from a given category), but the effect is certainly not just due to comprehending the words produced by the partner.

Interestingly, Wudarczyk et al. (2021) suggested that co-representation of a partners' words is specific to interacting with a human partner: They had participants complete the naming task with a

humanoid robot that named pictures while standing next to them (both the participant and the robot wore noise-cancelling headphones). When sharing the task with the robot, there was no evidence for co-representation of words (i.e., no increased semantic inhibition). However, participants were facilitated in naming semantic categories that were shared with the robot (compared to categories they named on their own). These findings suggest that the robot's task affected how quickly participants accessed meaning but not the downstream process of lexicalization.

In conclusion, there is good evidence that speakers co-represent whether others are speaking or about to speak, and that such co-representation affects concurrent language production, thus suggesting an involvement of production processes in the representation of others' utterances. It is less clear whether speakers co-represent the content of others' utterances in detail, but if they do, it is more likely when they are not preparing to speak at the same time. When speakers are preparing to speak, increasing the saliency of the production partner or making coordination with the partner a task requirement do not seem to make speakers more likely to engage in detailed co-representation. However, there is some indication that partner presence may increase the strength of co-representation for participants who are not preparing to speak at the same time (compare Kuhlen & Abdel-Rahman, 2017, with Kuhlen & Abdel-Rahman, 2021).

2. Joint Stroop and joint Picture-Word Interference tasks

We now review studies in which a picture-word interference (PWI) task or a Stroop task was split across two participants. In both tasks, participants produce a verbal response (picture name or ink color) while ignoring an irrelevant written word (distractor or color word); ignoring this information is hard and interferes with participants' responses. In the Stroop task, incongruent trials (where ink color and color word mismatch) lead to more errors and longer response times than congruent trials (MacLeod, 1991); in PWI, semantically related distractor words typically lead to more errors and longer naming times than semantically unrelated words (Glaser & Düngelhoff, 1984). In the joint version of these tasks, irrelevant information is associated with a partner's response and the question is how this association affects the level of interference.

The evidence from joint spatial action tasks (e.g., Simon task, SNARC task; Knoblich et al., 2011) is that participants represent their partner's response as well as their own, so when a task-irrelevant feature of the stimulus evokes the partner's response, they respond more slowly because of interference between their own and their partner's response. For example, Sebanz et al. (2003) had participants perform a spatial compatibility ("Simon") task. Participants saw a finger wearing a red or green ring that pointed left or right. When a single participant responded to red stimuli by pressing a button on the left and to green stimuli by pressing a button on the right, they produced faster responses when the finger pointed toward the button that they had to press than when it pointed to the other button. When participants responded in pairs, if one participant responded to (say) red stimuli and the other participant did not respond, there was no spatial compatibility effect (individual task). But when one participant responded to red stimuli and the other responded to green stimuli, the compatibility effect returned (joint task).

In contrast to this evidence for a "joint" spatial compatibility effect, Saunders and colleagues found no evidence for utterance co-representation in a joint version of the Stroop task (Saunders et al., 2019). In this button-press study, one participant responded to (say) yellow and blue words and the other responded to red and green words. They reasoned that, if the partner's response is represented, then there should be similar interference on incongruent trials where the written word corresponds to the alternative colour assigned to the same participant (own-colour trials; e.g., the word *yellow* written in blue) or one of the two colours assigned to the partner (other-colour trials; e.g., the word *red* written in blue), and there should be greater interference on both own-colour and other-colour trials than when the written word does not correspond to any response alternative (neutral trials; e.g., the word *purple* written in blue). Both of these predictions were supported, but critically a similar pattern was also found in the individual version of the task, when a single participant responded to two colours only, suggesting the results were not due to co-representation of the partner's utterances.

In a joint version of the verbal Stroop task, Pickering et al. (2021) also found no evidence that interference increased in the joint compared to the individual version. In the joint experiment, participants responded to words appearing in one ink colour, while their partners responded to words appearing in the other ink colour; in the individual experiment, participants had the same task but

partners did not respond to the other ink colour. Crucially, however, interference was greater in the joint than in the individual version when participants were additionally asked to monitor their partner's utterances for correctness (i.e., when they had to provide feedback). These findings suggest that the need to monitor another speaker's utterances may encourage a deeper representation of those utterances using the production system, thus leading to increased interference with production of one's own utterances. Accordingly, when participants' EEG was recorded while performing the joint and individual tasks with feedback (Demiral et al., 2016), the centro-parietal P3 (P3b) component was larger in the joint than the individual task on trials when it was the partner's turn to respond (i.e., silent trials for the participant), which suggests that participants represented their partner's upcoming response on these trials.

Interestingly, Demiral et al. (2016) also found a *reduced* congruency effect on the N2 component, which indexes perceptual conflict (Donkers & Van Boxtel, 2004), when participants responded in the joint compared to the individual task. This finding suggests that representing a co-actor's utterance may not only cause additional interference between competing response alternatives, but also attenuate perceptual conflict. Consistent with this hypothesis, a study that compared a joint and an individual version of the picture-word interference task (Sellaro et al., 2020) found a *reduced* semantic interference effect in a condition in which participants named pictures and were (falsely) told they had a partner in another room who read the superimposed distractor words. However, this reduction in the magnitude of the semantic interference effect only occurred when the distractor words were presented in case-alternating font, suggesting that representation of the partner's utterances can sometimes help participants ignore distracting information that is task-relevant for the partner, but only when processing of that information is made less automatic (in contrast, when the words were presented in regular font, or when participants believed their partner was naming the color of the pictures, comparable levels of interference were found in joint and individual versions). Finally, Kuhlen and Abdel Rahman (2022) found that when the PWI task is embedded in a communicative game, with one participant naming the distractor words and the other, co-present, participant naming the pictures, semantic interference is also greatly reduced (compared to a non-communicative, standard version of the PWI task). They suggested that naming pictures in a communicative setting may enhance semantic

facilitation at the conceptual level (due to distractor and target belonging to the same semantic category).

In sum, there is some suggestive evidence that in joint Stroop and PWI tasks participants co-represent a partner's response, even though this response is associated with stimulus features that are irrelevant for the participants' own task. Specifically, evidence from these tasks suggests that co-representation can alter the way in which task-irrelevant but partner-related information is processed (perhaps reducing perceptual conflict, or facilitating conceptual processing) and the associated response is selected (perhaps increasing response conflict). There is also some evidence that co-representation effects emerge when there is an explicit requirement to monitor the others' utterances or there is a clear communicative goal to the task. However, it is unclear precisely to what extent these effects can be attributed to co-representation of a partner's response as opposed to default processing of the irrelevant information. Moreover, these findings do not demonstrate that the production system is involved in co-representation, though they are consistent with others' utterances being represented similarly to one's own utterances.

3. Joint switching tasks

When unbalanced bilinguals are cued to switch languages while naming pictures, they sometimes (but not consistently; see Gade et al., 2021, for a recent meta-analysis) experience an asymmetrical cost – greater when switching from their second language (L2) into their first language (L1) than vice versa. This cost is thought to index the extent to which bilingual speakers need to inhibit the L1 in order to select an L2 word (e.g., Meuter & Allport, 1999). A few studies have asked whether switch costs are present when switching between speakers (i.e., from comprehension to production) and if so, whether they are also asymmetrical. Such findings would suggest that the language chosen by another speaker needs to be inhibited similarly to the language chosen by the speaker.

Two studies that directly compared switch costs from comprehension to production to switch costs within production found greater costs within production (Liu et al., 2020; Liu et al., 2021b), but they used recordings for the comprehension trials, so no partner was present (or assumed). In contrast, another series of studies used a joint language switching task with two co-present participants whose

EEG was recorded (Liu et al., 2021a; Liu et al., 2018; Liu et al., 2019; Xie et al., 2019; Zhang et al., 2019). Overall, behavioural measures revealed comparable switch costs within and between speakers. But analyses of the EEG data mostly revealed the asymmetric pattern – suggestive of increased inhibition when switching into L1 than into L2 – within but not between speakers. In fact, EEG markers of inhibition suggested that increased inhibition was applied to all between-person trials (relative to within-person trials), regardless of whether they included a language switch or not.

In comparison to cued switching, when unbalanced bilinguals switch language voluntarily in production, switch costs can be reduced, but typically they are not eliminated (e.g., de Bruin et al., 2018). In one study, unbalanced bilinguals switched between their languages voluntarily, while they took turns naming pictures themselves and listening to recordings of another speaker naming pictures (Liu et al., 2021b). While there were within-speaker switch costs, there were no switch costs from comprehension (of the recording) to production, and speakers were more likely to repeat the language they had previously used themselves than the language they had previously heard. But again, both these findings could be due to the absence of a partner. In Gambi and Hartsuiker (2016), two bilinguals took turns naming pictures in a joint task; one of them could voluntarily switch to the L2, while the other named pictures exclusively in the L1. The non-switching bilingual experienced switch costs from comprehending L2 words produced by their partner. Furthermore, individual pictures were named more slowly in L1 by the non-switching bilingual when their partner had previously named them in the L2 compared to when the partner had named them in the L1, showing that one speaker's language choices affected lexical retrieval within the other speaker's production system. However, this study did not include a comparison to within-speaker switching, so it does not address the question of whether production processes are involved in representing the switching partner's utterances.

In sum, there is some evidence for between-speakers switch costs both from cued and voluntary joint language switching paradigms. But more research is needed to clarify the extent to which others' utterances affect language selection mechanisms in production similarly to one's own previously produced utterances: A single study (Gambi & Hartsuiker, 2016) examined between-speaker costs in voluntary language switching in a truly joint task, and the evidence for cued joint language switching is hard to interpret because the behavioural and EEG findings diverge.

4. Choric speech and joint sentence production

Choric speech – speech produced synchronously with one or more other speakers – is relatively common, for example when people chant, pray, or protest. A few studies have examined the effect of asking speakers to speak synchronously with another speaker versus on their own. Cummins found that speakers are able to synchronize their speech with that of another speaker who is reading the same paragraph of text (Cummins, 2002, 2003, 2009). This synchronization is remarkably precise (discrepancies of only 40 ms), even without much practice; good synchronization is possible both with a "live" speaker and with a recording, but better with a live speaker (who can adapt as well) and with previous knowledge of the text.

Interestingly, choric speech tends to be slower than individual speech, and it tends to be less variable in terms of fundamental frequency, amplitude, and vowel duration (Poore & Ferguson, 2008), which may also facilitate synchronization. Moreover, in a fMRI study, Jasmin et al. (2016) asked participants to speak (1) individually (i.e., as baseline), (2) in sync with a recording, (3) at the same time as another live speaker but not in sync (i.e., with the live speaker producing a different utterance), or (4) in synch with a live speaker (who was producing the same utterance). Typically, speakers' responses to concurrent speech-like sounds in auditory areas are suppressed (so called, speech-induced suppression; e.g., Chang et al., 2013). Jasmin et al. (2016) observed speech-induced suppression for (2) and (3) but not for (4), suggesting that choric speech may be processed like other-produced speech.

In joint sentence production studies, participants produce short sentences with a confederate or another participant, with the constraint that the two speakers alternate and produce one word per turn. To our knowledge, three studies have used this paradigm (Fjaellingsdal et al., 2020; Himberg et al., 2015; Lelonkiewicz & Gambi, 2020). Himberg et al. (2015) showed that speakers entrained to each other's speech rhythm, and Fjaellingsdal et al. (2020) showed that turns were delayed after unexpected words. Taken together, these two studies demonstrate that speakers are able to perform this rather constrained joint production task by carefully monitoring and adapting to their partner's utterances. However, they did not include an individual condition, so it is hard to draw conclusions about the extent to which joint production resembles individual production on the basis of these two studies.

In contrast, Lelonkiewicz and Gambi (2020) asked two participants to type definitions for common English words, either on their own (individual task) or interacting with a naïve partner (joint task), and measured the timing and predictability of the resulting definitions. Consistent with findings from choric speech, interacting participants produced words with less variable delays than individuals; however, the duration of turns was not less variable in the joint than the individual condition, and jointly produced definitions were less predictable than definitions produced by individuals. In sum, there was some evidence for a reduction in variability during joint sentence production, which might help coordination, but it was not consistent across all measures. Overall, choric speech and joint sentence production studies highlight the fact that joint production may involve processes of adaptation to the other speaker or to the joint nature of the task that are absent from individual production.

Discussion

The study of joint production differs both from traditional monologic psycholinguistics and from the study of dialogue. It differs from traditional psycholinguistics because in joint language production tasks speakers do not believe they are producing language in isolation, but instead with a real or assumed partner who is also producing language either concurrently or in turn with the participant. It differs from the study of dialogue because the communicative and interactive aspects of language use in dialogue are typically stripped away or reduced to a bare minimum, in order to achieve greater experimental control and yield measures of performance that can be compared across joint and individual versions of the same task. In this way, researchers have been able to ask to what extent and how the mechanisms of language production – as reflected in traditional psycholinguistic tasks such as picture naming – are affected by representing others' utterances. Overall, the evidence suggests that such co-representation does take place, even when it is not task relevant. Moreover, it appears to use language production mechanisms, but typically in a manner that lacks detail: Speakers may represent whether their partner is preparing to speak, but not what they are about to say.

Our review highlighted three open issues. First, there is a question about the degree to which co-representation via the production system depends on the situation being a joint task or being perceived as a joint task by the speakers. Some findings suggest that co-representation may be stronger

when speakers are co-present (Kuhlen & Abdel Rahman, 2021) and when they are explicitly asked to monitor each other's utterances (Pickering et al., 2021), but other studies have found limited evidence for co-representation even though they explicitly asked speakers to coordinate their utterances (Hoedemaker & Meyer, 2018). A second open question is the extent to which co-representation is cognitively demanding and, relatedly, whether speaking and co-representing simultaneously reduce the resources available for co-representation (compared to co-representing without a simultaneous language production task) and therefore make such representations fairly undetailed. While our review generally supports this claim, no study has yet provided a direct comparison between co-representation when the participant is simultaneously speaking/preparing to speak and when they are not. Finally, choric speech and joint sentence production studies suggest there may be processes that are unique to joint production (adaptation, variability reduction), but these have not been extensively investigated and it is unclear how they may affect the comparison between joint and individual versions of the same task.

In sum, the evidence from join language production tasks lends some support to simulation-based theories (Dell & Chang, 2014; Pickering & Garrod, 2013; Pickering & Gambi, 2018) – that is, to the hypothesis that speakers can use their own language production system to represent the utterances of other speakers. However, this body of evidence also makes clear that simulation of others' speakers typically stops well short of full lexical access, and that the effects of co-presentation on language production do not always parallel the effects of previous production – suggesting that others' utterances are often represented differently from our own previous utterances. This has important implications for theory development, and future research should systematically investigate which factors influence the flexible use of simulation-based mechanisms (e.g., nature of the communicative context, partner's identity, cognitive load).

One possibility, based on the evidence reviewed above, is that comprehenders engage their production system by simulating what they encounter (in part to support prediction; Pickering & Gambi, 2018). The extent to which they simulate may depend on the situation. For example, when their task is to produce one utterance, they may inhibit their simulation of another utterance – accounting for the undetailed nature of co-representation when participants are concurrently engaged in a language

production task. Such inhibition of detailed co-representation may be benefical in communicative contexts, where it may help facilitate conceptual processing (see Kuhlen & Abdel Rahman, 2022). Additionally, they may be more likely to engage in simulation of partners that are more human-like (as is the case for simulation on non-verbal actions; e.g., Tsai & Brass, 2007).

Given that the communicative/interactive aspects of joint language use in conversation are intentionally stripped away in these tasks, one might ask whether the findings we have reviewed in this chapter bear relevance to understanding the processes that support successful between-speaker coordination in naturalistic conversations. We argue that they do, for two reasons. First, these tasks allow us to isolate the effect of others' utterances on production above and beyond the known effects of comprehension on production (e.g., priming). In addition, they make it possible to test the involvement of production processes in the representation of others' utterances, because they allow a direct comparison between the effect of comprehension on production and the effect of production on subsequent production. In sum, this body of work contributes to the theoretical understanding of the mechanisms of other representation in language production, which is relevant to dialogue as well as monologue.

References

Baus, C., Sebanz, N., de la Fuente, V., Branzi, F. M., Martin, C., & Costa, A. (2014). On predicting others' words: Electrophysiological evidence of prediction in speech production. *Cognition, 133*(2), 395-407.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology, 79*(6), 941-952.

Belke, E. (2013). Long-lasting inhibitory semantic context effects on object naming are necessarily conceptually mediated: Implications for models of lexical-semantic encoding. *Journal of Memory and Language, 69*(3), 228-256.

Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition, 104*(3), 437-458.

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.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*(2), B13-B25.

Brehm, L., Taschenberger, L., & Meyer, A. (2019). Mental representations of partner task cause interference in picture naming. *Acta Psychologica, 199*, 102888.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1482-1493.

Brown, A. S. (1981). Inhibition in cued retrieval. *Journal of Experimental Psychology: Human Learning and Memory, 7*(3), 204-215.

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences, 110*(7), 2653-2658.

Corps, R., 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.

Crocker, M. W., Pickering, M., & Clifton, C. (Eds.) (2000). *Architectures and mechanisms for language processing*. Cambridge University Press.

Cummins, F. (2002). On synchronous speech. *Acoustics Research Letters Online, 3*(1), 7-11.

Cummins, F. (2003). Practice and performance in speech produced synchronously. *Journal of Phonetics, 31*(2), 139-148.

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics, 37*(1), 16-28.

de Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2018). Voluntary language switching: When and why do bilinguals switch between their languages? *Journal of Memory and Language, 103*, 28-43.

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634), 20120394.

Demiral, Ş. B., Gambi, C., Nieuwland, M. S., & Pickering, M. J. (2016). Neural correlates of verbal joint action: ERPs reveal common perception and action systems in a shared-Stroop task. *Brain Research, 1649*, 79-89.

Donkers, F. C., & Van Boxtel, G. J. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain and Cognition, 56(2)*, 165-176.

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*, 5460.

Gade, M., Declerck, M., Philipp, A. M., Rey-Mermet, A., & Koch, I. (2021). Assessing the evidence for asymmetrical switch costs and reversed language dominance effects – A meta-analysis. *Journal of Cognition*, 4(1): 55, pp. 1–32. DOI: https://doi.org/10.5334/ joc.186

Gambi, C., Cop, U., & Pickering, M. J. (2015a). How do speakers coordinate? Evidence for prediction in a joint word-replacement task. *Cortex, 68*, 111-128.

Gambi, C., & Hartsuiker, R. J. (2016). If you stay, it might be easier: Switch costs from comprehension to production in a joint switching task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(4), 608-626.

Gambi, C., & Pickering, M. J. (2017). Models linking production and comprehension. In E. M. Fernández & H. Smith Cairns (Eds.), *Handbook of Psycholinguistics* (pp. 157-181). Hoboken, NJ: John Wiley & Sons Inc.

Gambi, C., Van de Cavey, J., & Pickering, M. J. (2015b). Interference in joint picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(1), 1-21.

Gambi, C., Van de Cavey, J., & Pickering, M. J. (2022). Representation of others' synchronous and asynchronous sentences interferes with sentence production. *Quartelrly Journal of Experimental Psychology*, https://doi.org/10.1177/17470218221080766 .

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*(2), 181-218.

Gauvin, H. S., & Hartsuiker, R. J. (2020). Towards a new model of verbal monitoring. *Journal of Cognition, 3*(1), 17. doi:10.5334/joc.81

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 640-654.

Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research, 34*(4), 365-399.

Hartsuiker, R. J., Catchpole, C. M., De Jong, N. H., & Pickering, M. J. (2008). Concurrent processing of words and their replacements during speech. *Cognition, 108*(3), 601-607.

Hickok, G. (2013). Predictive coding? Yes, but from what source?. *Behavioral and Brain Sciences, 36*(4), 358. doi:10.1017/S0140525X12002750

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.

Hoedemaker, R. S., Ernst, J., Meyer, A. S., & Belke, E. (2017). Language production in a shared task: Cumulative semantic interference from self-and other-produced context words. *Acta Psychologica, 172*, 55-63.

Hoedemaker, R. S., & Meyer, A. S. (2018). Planning and coordination of utterances in a joint naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(4), 732–752. doi:10.1037/xlm0000603.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition, 100*(3), 464-482.

Ivanova, I., Horton, W. S., Swets, B., Kleinman, D., & Ferreira, V. S. (2020). Structural alignment in dialogue and monologue (and what attention may have to do with it). *Journal of Memory and Language, 110*, 104052.

Jacobs, C. L., Cho, S. J., & Watson, D. G. (2019). Self-priming in production: Evidence for a hybrid model of syntactic priming. *Cognitive Science, 43*(7), e12749.

Jasmin, K. M., McGettigan, C., Agnew, Z. K., Lavan, N., Josephs, O., Cummins, F., & Scott, S. K. (2016). Cohesion and joint speech: right hemisphere contributions to synchronized vocal production. *Journal of Neuroscience, 36*(17), 4669-4680.

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* (Vol. 54, pp. 59-101). Burlington: Academic Press.

Kootstra, G. J., Van Hell, J. G., & Dijkstra, T. (2010). Syntactic alignment and shared word order in code-switched sentence production: Evidence from bilingual monologue and dialogue. *Journal of Memory and Language, 63*(2), 210-231.

Kuhlen, A., & Abdel Rahman, R. (2017). Having a task partner affects lexical retrieval: Spoken word production in shared task settings. *Cognition, 166*, 94-106.

Kuhlen, A., & Abdel Rahman, R. (2021). Joint language production: an electrophysiological investigation of simulated lexical access on behalf of task partner. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/xlm0001025

Kuhlen, A., & Abdel Rahman, R. (2022). Mental chronometry of speaking in dialogue: Semantic interference turns into facilitation. *Cognition, 219*, 104962.

Lelonkiewicz, J. R., & Gambi, C. (2020). Making oneself predictable in linguistic interactions. *Acta Psychologica, 209*, 103125.

Liu, H., Kong, C., de Bruin, A., Wu, J., & He, Y. (2020). Interactive influence of self and other language behaviors: Evidence from switching between bilingual production and comprehension. *Human Brain Mapping, 41*(13), 3720-3736.

Liu, H., Li, B., Wang, X., & He, Y. (2021a). Role of joint language control during cross-language communication: evidence from cross-frequency coupling. *Cognitive Neurodynamics, 15*, 191-205.

Liu, H., Li, W., de Bruin, A., & He, Y. (2021b). Should I focus on self-language actions or should I follow others? Cross-language interference effects in voluntary and cued language switching. *Acta Psychologica, 216*, 103308.

Liu, H., Xie, N., Zhang, M., Gao, X., Dunlap, S., & Chen, B. (2018). The electrophysiological mechanism of joint language switching: evidence from simultaneous production and comprehension. *Journal of Neurolinguistics, 45*, 45-59.

Liu, H., Zhang, M., Pérez, A., Xie, N., Li, B., & Liu, Q. (2019). Role of language control during interbrain phase synchronization of cross-language communication. *Neuropsychologia, 131*, 316-324.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin, 109*(2), 163-203.

Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*(1), 25-40.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition, 114*(2), 227-252.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological Bulletin, 134(3),* 427-459, https://doi.org/10.1037/0033-2909.134.3.427.

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. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329-392. doi:http://dx.doi.org/10.1017/S0140525X12001495

Pickering, M. J., & Garrod, S. (2021). *Understanding dialogue: Language use and social interaction*. Cambridge, UK: Cambridge University Press.

Pickering, M. J., McLean, J. F., & Gambi, C. (2021). *Interference in the Shared-Stroop task: A comparison of self- and other-monitoring*. Manuscript submitted for publication.

Poore, M. A., & Ferguson, S. H. (2008). Methodological variables in choral reading. *Clinical Linguistics & Phonetics, 22*(1), 13-24.

Reitter, D., Moore, J. D., & Keller, F. (2006). *Priming of syntactic rules in task-oriented dialogue and spontaneous conversation.* Paper presented at the 28th Annual Conference of the Cognitive Science Society, Vancouver, BC, Canada.

Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology, 2*. doi:10.3389/fpsyg.2011.00307

Saunders, D. R., Melcher, D., & van Zoest, W. (2019). No evidence of task co-representation in a joint Stroop task. *Psychological Research, 83*(5), 852-862.

Schoot, L., Hagoort, P., & Segaert, K. (2019). Stronger syntactic alignment in the presence of an interlocutor. *Frontiers in Psychology, 10*, 685.

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition, 88*(3), B11-B21.

Segaert, K., Kempen, G., Petersson, K. M., & Hagoort, P. (2013). Syntactic priming and the lexical boost effect during sentence production and sentence comprehension: An fMRI study. *Brain and Language, 124*(2), 174-183.

Sellaro, R., Treccani, B., & Cubelli, R. (2020). When task sharing reduces interference: Evidence for division-of-labour in Stroop-like tasks. *Psychological Research, 84*(2), 327-342.

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

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex, 20*(4), 912-928.

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.

Tsai, C. C., & Brass, M. (2007). Does the human motor system simulate Pinocchio's actions? Coacting with a human hand versus a wooden hand in a dyadic interaction. *Psychological Science, 18(12),* 1058-1062.

Tydgat, I., Stevens, M., Hartsuiker, R. J., & Pickering, M. J. (2011). Deciding where to stop speaking. *Journal of Memory and Language, 64*(4), 359-380.

Wudarczyk, O. A., Kirtay, M., Pischedda, D., Hafner, V. V., Haynes, J. D., Kuhlen, A. K., & Abdel Rahman, R. (2021). Robots facilitate human language production. *Scientific Reports, 11, 16737, https://doi.org/10.1038/s41598-021-95645-9*

Xie, N., Li, B., Zhang, M., & Liu, H. (2019). Role of top-down language control in bilingual production and comprehension: Evidence from induced oscillations. *International Journal of Bilingualism, 23*(5), 1041-1063.

Zhang, M., Wang, X., Wang, F., & Liu, H. (2019). Effect of cognitive style on language control during joint language switching: An ERP study. *Journal of Psycholinguistic Research, 49*(3), 383-400.