Cognitive, Affective, & Behavioral Neuroscience
https://doi.org/10.3758/s13415-023-01103-6

SPECIAL ISSUE/UNCERTAINTY

# Utterance planning under message uncertainty: evidence from a novel picture-naming paradigm

**Arella E. Gussow**[1] · **Maryellen C. MacDonald**[1]

Accepted: 14 April 2023
© The Psychonomic Society, Inc. 2023

## Abstract

Language researchers view utterance planning as implicit decision-making: producers must choose the words, sentence structures, and various other linguistic features to communicate their message. To date, much of the research on utterance planning has focused on situations in which the speaker knows the full message to convey. Less is known about circumstances in which speakers begin utterance planning before they are certain about their message. In three picture-naming experiments, we used a novel paradigm to examine how speakers plan utterances before a full message is known. In Experiments 1 and 2, participants viewed displays showing two pairs of objects, followed by a cue to name one pair. In an Overlap condition, one object appeared in both pairs, providing early information about one of the objects to name. In a Different condition, there was no object overlap. Across both spoken and typed responses, participants tended to name the overlapping target first in the Overlap condition, with shorter initiation latencies compared with other utterances. Experiment 3 used a semantically constraining question to provide early information about the upcoming targets, and participants tended to name the more likely target first in their response. These results suggest that in situations of uncertainty, producers choose word orders that allow them to begin early planning. Producers prioritize message components that are certain to be needed and continue planning the rest when more information becomes available. Given similarities to planning strategies for other goal-directed behaviors, we suggest continuity between decision-making processes in language and other cognitive domains.

**Keywords** Language production · Utterance plan · Message uncertainty · Word-order choice

## Introduction

"What should we have for dinner?" your partner asks, as you are surveying the contents of the refrigerator. This situation puts you in one obvious type of decision making: settling on a dinner plan, which is an explicit decision that might even involve conscious deliberation between meal options. However, the situation also contains several less obvious types of implicit decision making involving language production, which are required for you to be able to respond.

The act of language production—speaking, writing, or using a sign language—often is construed as requiring extensive implicit decision making, because producers must choose words, word orders, and other language features to suit the message they wish to convey (Anders et al., 2015; Bock, 1995; Garrett, 1989; Gennari et al., 2012; Koranda et al., 2020, 2022; Levelt, 1989; MacDonald, 2013). The dinner conversation above also creates an additional challenge of decision making under *message uncertainty*, in that the conventions of social interaction might compel you to reply before the original decision—what to have for dinner—is known. If you do not fully know the answer yet but want to reply, how do the decision processes of language production operate?

This article investigates language production and its associated decision processes under message uncertainty. We report studies using a novel paradigm in which participants produce language to describe pictures, in situations in which they are either certain or temporarily uncertain about what is to be described. We uncover interesting differences in production decisions as a function of uncertainty; but beyond our particular results, we aim to build bridges between the fields of language production and decision making (Anders et al., 2015; Koranda et al., 2022). We first review some of the decision making components in language production and then consider language production under message

✉ Arella E. Gussow
gussow@wisc.edu

[1] Department of Psychology, University of Wisconsin-Madison, 1202 West Johnson St, Madison, WI 53706, USA

Published online: 15 May 2023

uncertainty, while also pointing to parallels with decision making in the motor action domain.

## Decision-making in language production

Language production is typically assumed to begin with a *message*, a package of information that the speaker would like to convey (Garrett, 1989; Levelt, 1989). Given this communicative goal, speakers must convert the message into overt behavior—an utterance—that others can comprehend. However, a given message can often be expressed in several different ways, and the speaker must decide between alternative utterance forms (Bock, 1982, 1995). Responding to the question about dinner above, for example, you might indicate a dinner preference by saying "Pasta," or "How about pasta?"; "Let's have pasta," "Pasta would be good," and so on.

Many language researchers view the process of utterance planning as continuous decision-making, as producers implicitly choose between various language options for conveying their message (Garrett, 1989; Gennari et al., 2012; Koranda et al., 2020, 2022; Levelt, 1989; MacDonald, 2013). Choice of one utterance form over others is affected by perceptual context (Gleitman et al., 2007), shared knowledge between speakers (Heller et al., 2009), frequency and priming of words (Bock, 1986; Branigan and McLean, 2016), or sentence structures (Bock, 1986), and several other domain-general cognitive constraints, including effort minimization (Koranda et al., 2022; MacDonald, 2013).

As with all forms of action, language production requires the development of an internal plan before execution (Lashley, 1951; MacDonald, 2016; Rosenbaum et al., 2007). Production decisions are required at several different planning levels, including selecting the words to include in the utterance, arranging them in a certain order, planning the words' pronunciation or spelling, and articulating the resulting utterance plan (Bock, 1995; Garrett, 1989; Levelt, 1989). These various planning stages frequently overlap in time, requiring rapid coordination of several different decisions (Harley, 1984; Smith and Wheeldon, 2004; Vigliocco and Hartsuiker, 2002).

To date, much of the research on utterance planning has focused on situations in which the speaker knows the full message to convey (Allum and Wheeldon, 2007; Barthel et al., 2016; Bock, 1995; Bögels et al., 2018; Griffin, 2001; Jongman et al., 2020; Smith and Wheeldon, 2004), such as in the dinner example, when you may have fully decided on pasta for dinner before answering your partner. Production models also have typically assumed that decisions about the message are completed before most other stages of utterance planning begin (Levelt, 1989). However, there is evidence that producers can update utterance plans during production if there is a change in the message (Brown-Schmidt and Konopka, 2008), suggesting some flexibility and overlap in the timing of message formulation, utterance planning, and articulation (see also Brown-Schmidt and Konopka, 2015; Ferreira and Swets, 2002). Moreover, research on turn-taking in conversation suggests that speaker A can begin planning their response to speaker B early, while speaker B is still speaking (Barthel et al., 2016; Bögels et al., 2015; Corps et al., 2018; Levinson, 2016). Because speaker A's response will depend on speaker B's ongoing utterance, such early response planning suggests that some degree of planning occurs under message uncertainty. In fact, message updating is perhaps more likely to occur in such interactive conversations, because the listener can give the speaker feedback cues (e.g., back-channel feedback, facial expressions, etc.) (Schegloff, 1982) that might cause the speaker to update their message (Tolins and Fox Tree, 2014).

This research casts doubt on the view that production processes must await a fully settled message, although there is little empirical work on this topic. Notably, similar ideas of interactivity and cascading activation have been discussed extensively, especially at the lexico-semantic level of word choice (Indefrey, 2011; Rapp and Goldrick, 2000; Roelofs, 2008). This leaves the question of how production proceeds under *message* uncertainty—given that even the higher-level communicative goal is not yet settled.

## Language and action under uncertainty

Some insight into production under uncertainty might be gained from the motor control literature, as the need to decide between alternative behaviors is not unique to language planning but rather a general property of goal-directed actions (Koranda et al., 2020; Lebkuecher et al., 2022; Wolpert and Landy, 2012). For example, theories of motor control suggest that probabilistic decision-making processes determine motor choices, such as which limb to use for an action, which path to follow to a target, or when to initiate movement (for a review, see Wolpert and Landy, 2012). The strategies used for action under uncertainty are known to be modulated by factors such as movement speed, target probabilities, target similarity, target proximity, and individual differences in the approach to tasks (Wong and Haith, 2017). Participants appear to weigh various sources of information—visual context, general knowledge, prior experiences—for statistically optimal decision-making about when and which action to execute (Battaglia and Schrater, 2007; Faisal and Wolpert, 2009).

In language, action, and other goal-directed behaviors, uncertainty often arises because of time constraints, forcing the actor to begin their action before they have the

full perceptual information needed to complete it (Faisal and Wolpert, 2009; Levinson, 2016). The actor therefore must decide how much information is sufficient for planning and initiating action. In sensorimotor tasks, previous studies have shown near optimal performance when there are necessary trade-offs between perception and motor action uncertainties (Battaglia and Schrater, 2007; Faisal and Wolpert, 2009). Strategic planning also appears in the face of goal uncertainty, such as aiming a movement in between two potential targets, and flexibly adapting the motor plan once a single target is disambiguated (Chapman et al., 2010; Gallivan et al., 2015; Krüger and Hermsdörfer, 2019). Given the similar strategic and flexible nature of language production (Ferreira and Swets, 2002; MacDonald, 2013), we also might expect speakers to use certain form flexibilities, such as word order, to facilitate planning under message uncertainty. This type of finding would suggest continuity between utterance planning under uncertainty and action planning more generally.

In the current study, we develop a new paradigm to investigate how speakers plan their utterances when they have only partial message information. Our paradigm is conceptually related to those in reaching studies in which it is temporarily unknown which of several locations will be the target of the reaching action (Chapman et al., 2010; Gallivan et al., 2015). The language implementation is very different from reaching tasks, however, and predictions also diverge. Whereas participants in reaching studies produce movement trajectories that are in between potential targets, such intermediate strategies do not work for word production—for someone who is uncertain whether to say "pasta" or "salad," saying a blend of the two words is not an effective strategy. Instead, we hypothesize that producers will rely on the known flexibility of sequencing in utterance planning. Specifically, we predict that speakers will begin planning components of the utterance that are certain to be useful, while the rest can be planned later, as more information becomes available. In our example above, if a speaker knows that they want to have salad with their dinner but have not yet decided on the main dish, they might say "salad as a side…salad and pasta?"—such that uttering the certain part (the salad) allows more time to plan the uncertain part (the pasta).

If so, this production strategy will affect speakers' word order, as they prioritize the components that are certain to be useful and choose to place them first in the utterance. Early planning should allow speakers to begin production sooner, showing a benefit in speech initiation latencies. Such findings would suggest that word order can be shaped by implicit planning strategies that maximize production efficiency, including early planning, even under message uncertainty.

## Experiment 1

Production experiments commonly provide participants with the message they need to convey, such as a scene to describe (Bock, 1996). While this method is helpful for experimental control, it is not well-suited for studying planning under message uncertainty. In the current study, we modify standard picture-description tasks to first display several pictures, and only later provide information about which two pictures are to be named. Two different types of display allow manipulation of message uncertainty: in one condition (Overlap), participants have early information about one picture to be named, whereas in the other condition (Different), this information is not available. Thus, our manipulation affects *how much* of the message information participants have at the start of the trial, and consequently, how much of the utterance they can plan. That is, in the Overlap condition, participants already know half of the message (one of the pictures to be named) from the beginning but are uncertain about the other, whereas in the Different condition, they see several options but are uncertain about all of them. By carefully controlling the timing of when the full information is revealed, we can examine how speakers incorporate any partial message information into their utterance planning. We predict that when early information is available, producers will begin to plan their utterances early, prioritizing the picture name that will likely be needed. This should result in that name being produced first, with shorter initiation latencies compared with the condition in which no prediction is possible.

### Method

#### Participants

Sixty-six undergraduate students, all native English speakers, participated for course credit. Sample size was determined based on pilot testing and a power analysis conducted using PANGEA (Westfall, 2016), aiming for at least 80% power with a moderate effect size, and allowing leeway in case of participant exclusions. Data from seven additional participants were excluded from analyses: four who did not follow instructions, two due to equipment failure, and one who was a nonnative English speaker.

#### Stimuli

Object images from the MultiPic database (n = 320, Duñabeitia et al., 2017) were used as stimuli, assigned to 80 displays of four objects each. Object names within a display were matched on syllable count and word frequency (Balota

et al., 2007). For each display, two objects were randomly chosen as the target pair, and the remaining two were the competitor pair. Objects were never repeated across displays.

Each display could appear in one of two conditions: 1) Overlap, where one of the target pictures was repeated in the display, replacing one of the competitors; 2) Different, where there was no overlap between targets and competitors. Each display contained one pair of images appearing on either side of a computer screen (Fig. 1). The two images within each pair rotated around each other to avoid position effects on naming order.

### Procedure

Participants came into the lab and were seated in a soundproof booth with a computer and a freestanding microphone. They first completed a familiarization phase, where they viewed each of the images with its name on screen and said the name aloud. Each image appeared for 3 seconds with a 1 second interstimulus fixation; the familiarization phase lasted approximately 15 minutes.

Participants then received the main task instructions, including example displays and four practice trials with feedback. In each trial, a display was presented, and after 2.2 seconds, a gray background appeared on one side of the screen, indicating the target pair. Participants' task was to answer the question, "Which are the target images?" They were told that another participant would later listen to their recordings to identify the targets. This cover story was added to encourage participants to treat the task as information sharing with another individual. Participants were instructed to produce a full phrase, such as "the vest and the pear" for the example in Fig. 1, rather than simply naming the objects. Participants were told that sometimes pictures would be duplicated in displays but that they should name the pair cued by the gray background. Participants were not told anything about uncertainty or early language planning and remained blind to the goal of the experiment.

The test phase of 80 trials lasted approximately 20 minutes. Participants were randomly assigned to one of eight lists, counterbalancing which picture was doubled, which competitor was replaced, and condition. Images were offset from the center at the start of the trial. The target side and the initial position of each image within the pair were pseudorandomized, i.e., appearing on each side of the screen (left vs. right) and in each position within the pair (top vs. bottom) in equal proportions across trials. Order of items (displays) was randomized per participant. After completing the experiment, participants answered a survey to provide feedback, including their guess about the experiment purpose and/or any strategies that they developed.

## Analyses and results

### Preprocessing

Speech files were manually transcribed. An automated script then coded each trial for accuracy in object naming. Object names that differed from the names learned in the familiarization phase were coded as inaccurate, even if the produced name was generally acceptable in English. Trials in which one or both objects were named inaccurately (25%) were removed from analyses, as were trials with disfluencies (3%).

The FAVE Program Suite (Rosenfelder et al., 2014) was used to extract onset and offset times of each word in each utterance. Before performing latency analyses, we further removed responses that did not follow the 5-word conjoined noun phrase structure (21%), trials where participants began speaking before the cue (6%), and responses with a total duration of more than 2.5 SD above the mean duration (3%). Note that percentages refer to the percentage of remaining observations after the previous step of data cleaning. This left 2,797 observations (1,419 Different, 1,378 Overlap; 53% of all observations) from 55 participants for each word position in the latency analyses.

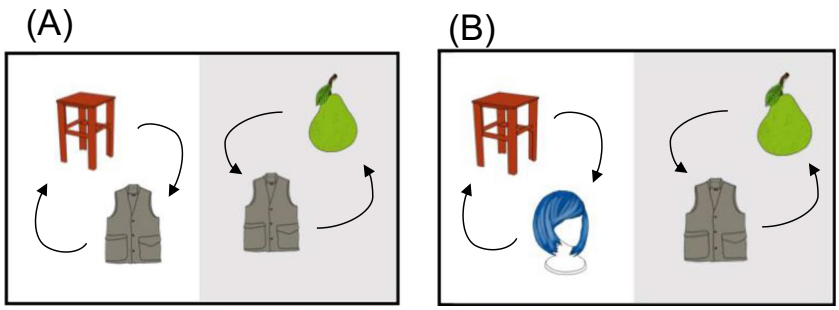


**Fig. 1** Visual displays in the Overlap (A) and Different (B) conditions. Every two images rotated around each other, as illustrated by the arrows. Arrows were not visible to participants. The gray background appeared after 2.2 seconds of exposure, indicating the targets

### Word order

We first tested whether participants were more likely to name the overlapping object first in their responses in the Overlap condition. We ran a mixed-effects logistic regression model regressing Order (nonoverlapping vs. overlapping object produced first) on the intercept, including by-participant and by-item random intercepts (the maximal random effects structure; Barr, 2013). The coefficient is reported in log-odds and represents the model intercept. The null model for comparison sets the intercept at zero, which in log-odds is equivalent to 50%. A positive significant intercept therefore indicates above-chance preference for one word order over the other. As expected, participants were significantly above chance in producing the overlapping object first in their utterance ($b = 0.75$, SE = 0.1, $z = 7.28$, $p < 0.0001$; Fig. 2).

### Latencies

We next examined speech initiation latencies (onsets) for each word in participants' utterances. Onsets were calculated as the time between the cue appearing and the first phoneme of each word and were log-transformed for analyses. The predictor variables were word Position (1-5), Condition (Different, Overlap; coded [0,1]), and Order of naming (nonoverlapping first, overlapping first; coded [0,1]). Order was nested within condition (Condition/Order), because it was only relevant in the Overlap condition. Position was coded by using successive difference contrasts, where each word position is compared to the preceding word position (2-1, 3-2, 4-3, 5-4). Each contrast therefore measures the difference between the onset of a given word and the onset of the prior word in that utterance.

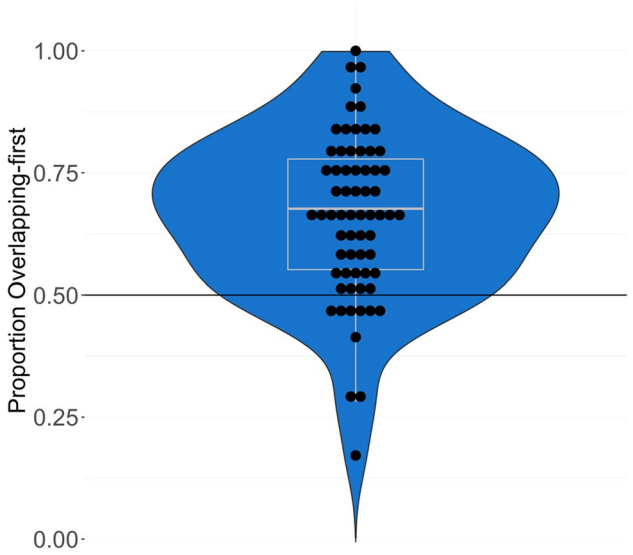


**Fig. 2** Violin plot of the proportion of trials in the Overlap condition in which participants named the overlapping target first (Experiment 1). A boxplot is overlayed in gray to identify quartiles and the median proportion. The horizontal black line identifies chance level. Points reflect individual participant means

We ran a mixed-effects linear regression model regressing Onset on Position, Condition/Order, and the interaction between Position and Condition/Order. Including the nested term (Condition/Order) is equivalent to including fixed effects of Condition and the interaction between Condition and Order. The maximal random effects structure to converge included by-subject and by-item random intercepts, by-subject random slopes for Condition/Order and for the interaction between Condition:Order:Position, and a by-item random slope for Condition/Order. Significance values are reported using Satterthwaite's method *t*-tests.

Results show a significant interaction between Condition and Order, $b = -0.01$, SE = 0.002, $t = -4.36$, $p < 0.0001$; speakers had shorter word onsets when the overlapping object was named first (Fig. 3). There was no significant effect of Condition ($b = -0.0003$, SE = 0.001, $t = -0.2$, $p = 0.8$), suggesting the advantage in onset times is specific to trials where the overlapping object was named first, and not an overall difference between Overlap and Different conditions. There also was no significant interaction between Condition and Position, or between Condition, Position, and Order ($ps > 0.1$), indicating the effect of Order was driven by differences in the onset of the first word, and not due to differing word durations or pauses throughout the utterance.

Responses from the post-experiment survey showed that none of the participants were able to guess the experiment purpose. When participants were specifically asked about strategies or attempts to plan early, 13 participants mentioned or alluded to the overlapping image as playing a part in their strategy. We reran the analyses excluding these participants; results held for both the order analysis ($b = 0.72$, $SE = 0.05$, $z = -13.46$, $p < 0.0001$) and the latencies (Condition by Order interaction: $b = -0.008$, SE = 0.002, $t = -2.98$, $p < 0.01$).

## Experiment 1 discussion

Results of Experiment 1 suggest that speakers prioritized planning words that they were certain to need (the overlapping object), placing them first in their utterance plan. This early planning based on partial message information provided a benefit in onset latencies, even though speakers only began articulation after the visual cue appeared and the full message was available.

As Figure 3 shows, latency differences emerged on the very first word: "the." This effect is interesting, because "the" is the first word to be said in every trial, independent of condition or order of the pictures described. The shorter latencies when the overlapping picture was said first suggests

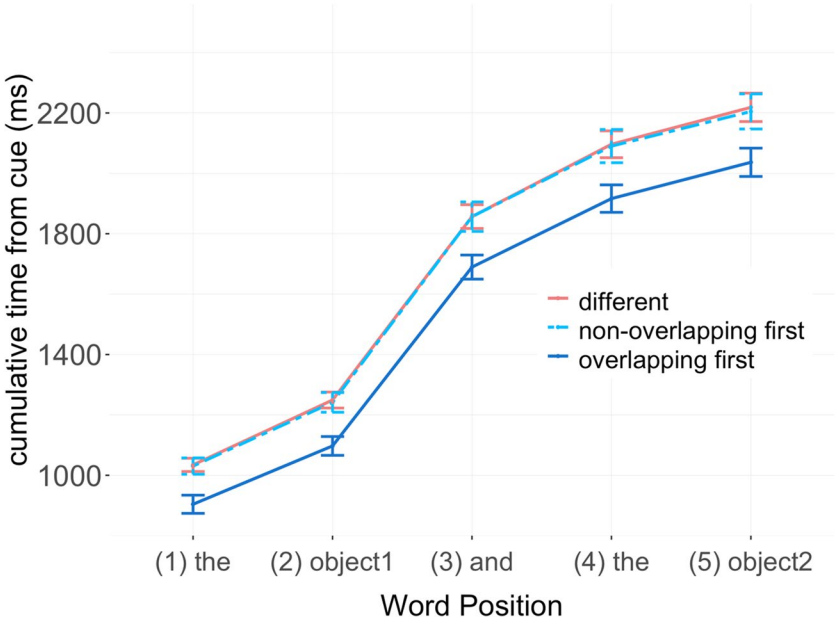


**Fig. 3** Model predictions for onsets in Experiment 1. Data from the Overlap condition are divided into trials where participants named the overlapping target first (dark blue) or the nonoverlapping target first (dashed, light blue). Error bars represent ±1 SE

that speakers do not begin to produce the first phrase (e.g., "the vest") until they have planned the noun component; that is, they do not immediately produce "the" and then pause until the noun ("vest") is ready. Although speakers can sometimes pause or extend the pronunciation of "the" to provide more planning time for a noun, as in "Have you seen theee, uh, remote?" (Arnold et al., 2003; Clark and Fox Tree, 2002), most production studies using pictured stimuli have found that speakers typically initiate production of phrases (e.g., "the vest") after the entire minimal phrase has been planned (Allum and Wheeldon, 2007). The data from Experiment 1 conform to that pattern.

With the goal of replicating and extending the results of Experiment 1, we next conducted Experiment 2 as an online study that uses typed responses instead of speech. Typing presents an interesting alternative to speech, because it is slower and more protracted, which could result in different planning strategies (Snyder and Logan, 2014). In our study, the slower time course of typing might affect participants' production decisions, because they have more time to plan upcoming components even after production has begun. This could provide more time to perceive the information needed for their utterance and/or to make production decisions while typing is ongoing. Moreover, differences in the effort and time needed for typing versus speaking might affect participants' production decisions because of the costs and benefits involved in beginning production early. For example, the increased effort and time needed for typing might encourage participants to begin production early, so they can keep up with time pressures, leading to more extreme production biases compared with speech. Alternatively, it might deter participants from beginning production early because of the risk of error; it is harder and more time-consuming to correct typed errors compared with spoken errors. Thus although our predictions for typing are more exploratory, extending our results from spoken to typed productions not only allows replicating the effect found in Experiment 1 but also could provide new insights into how decision making in language planning compares across modalities.

## Experiment 2

To convert the task in Experiment 1 to an online format with typed responses, changes to the design and materials were needed. To limit the experiment length, we eliminated the familiarization phase and reduced the number of trials to 62 compared with 80 in Experiment 1. In creating the new set of items, we prioritized images that had shown high naming accuracy in Experiment 1.

We expected that participants would again prioritize more certain information and place overlap pictures first in their responses. Our predictions for latencies were less clear, given that typing is less practiced than speaking and likely to yield more variable data.

## Method

### Participants

Eighty-two undergraduate students, all native English speakers, participated for course credit. Given the online format of the study and concerns about participants' attention, we aimed for a larger sample size than Experiment 1. Data from 25 additional participants were excluded from analyses: four who reported technical difficulties, four due to a script error, eight nonnative English speakers, three who guessed the experiment goal, and six who did not complete all trials.

### Procedure

The experiment was coded in jsPsych (de Leeuw, 2015) and run online, via a lab server. Trial procedure was identical to Experiment 1, except that once the gray background appeared, a text box appeared at the bottom center of the screen. Participants could then type in their responses using any of the letter keys, the spacebar, and the delete key. The experiment lasted approximately 20 minutes.

## Analyses and results

### Preprocessing

The experiment script logged each keypress and its time stamp relative to the cue appearance. A pre-processing script then parsed the key strings into words, checked naming accuracy, and extracted time stamps for the first and last keys of each word. Trials were manually coded if participants used the delete key, misspelled a word, or did not use the instructed 5-word structure (43%).

Trials in which one or both objects were named inaccurately were removed from analyses (22%); misspellings were retained. For latency analyses, responses that used a different sentence structure than instructed (18%), and responses with a total duration of more than 2.5 SD above the mean duration (2%) were removed. In cases of misspellings or use of the delete key (38%), we analyzed latencies only for the first word in the sentence, which was always the word "the." Note that percentages refer to the percentage of remaining observations after the prior step of data cleaning. This left 3,198 observations (1,579 Different, 1,619 Overlap; 63% of all observations) from 68 participants for latency analyses of the first word, and 1,966 observations for each word in positions two through five (976 Different, 990 Overlap). Analyses were identical to Experiment 1, except that in the latency model we removed the by-item random intercept to allow convergence.

### Results

As in Experiment 1, participants were significantly above chance in naming the overlapping object first ($b = 0.78$, $SE = 0.1$, $z = 8.06$, $p < 0.0001$; Fig. 4).

Latency analyses showed a significant interaction between Condition and Order, indicating shorter onset latencies when the overlapping object was produced first in the typed response, $b = -0.006$, SE = 0.002, $t = -2.94$, $p < 0.01$ (Fig. 5). There was no significant effect of Condition ($p > 0.1$), but there was a marginally significant three-way interaction between Condition, Order, and the contrast of Positions [3-2], $b = 0.004$, SE = 0.002, $t = 1.68$, $p = 0.09$. Given the small number of observations available for word positions 2 to 5, especially for the rare nonoverlapping first responses, we are wary of interpreting interactions involving positions 2 to 5. Future studies might explore late-position latencies further, as position interactions could suggest differences in planning during production itself. In either case, initiation latencies were shorter for overlapping-first sentences, resembling the Experiment 1 results. Moreover, when excluding 22 participants who alluded in the post-experiment survey to any strategy involving the overlapping image, results held for the order analysis ($b = 0.61$, $SE = 0.1$, $z = 5.86$, $p < 0.0001$), and the latencies (Condition by

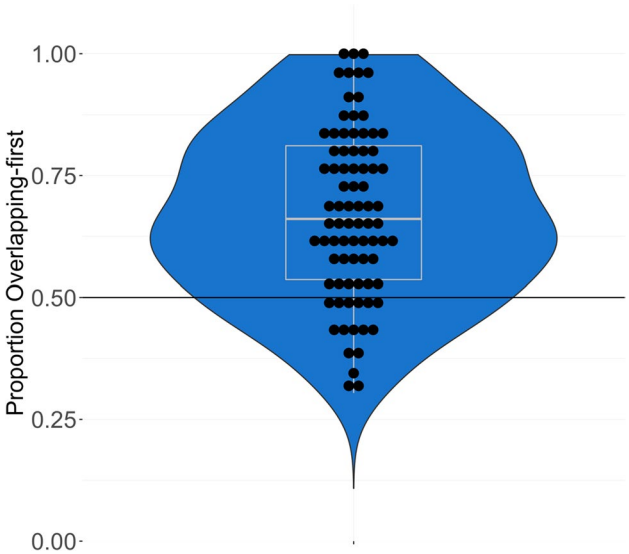


**Fig. 4** Violin plot of the proportion of trials in the Overlap condition in which participants typed the overlapping target first (Experiment 2). A boxplot is overlayed in gray to identify quartiles and the median proportion. The horizontal black line identifies chance level. Points reflect individual participant means

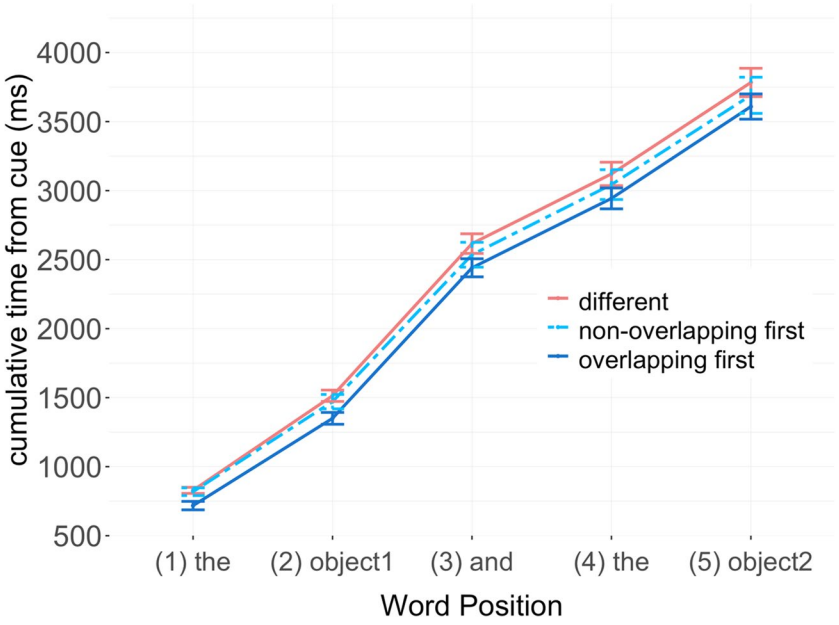


**Fig. 5** Model predictions for onsets in Experiment 2. Data from the Overlap condition are divided into trials where participants named the overlapping target first (dark blue) or the nonoverlapping target first (dashed, light blue). Error bars represent ±1 SE

Order interaction: $b = -0.006$, SE = 0.002, $t = -2.63$, $p < 0.05$, n = 52).

## Experiment 2 discussion

Results from Experiment 2 largely replicated those of Experiment 1: participants were more likely to name the overlapping object first, and this provided a benefit in typing initiation latencies. Despite numerous differences between speech and typing that could result in different strategies for production planning (Snyder and Logan, 2014), we find similar word order biases across modalities, indicating a robust planning bias. The converging findings support our hypothesis that producers can use partial message information to begin planning their responses early, with implications for word order and response times.

One question about our findings is whether production of the overlapping object first reflects conscious strategizing and/or implicit production decisions. While conscious strategizing certainly exists in everyday language use, we do not believe that it has a major role in our findings. First, when participants were probed about strategies used to begin planning their responses early, most did not report consciously planning the overlapping object first, and our results held even when excluding those that did. This is not surprising, given that evidence of early planning is found in natural conversation (Bögels, 2020), and is therefore unlikely to implicate deliberate strategies for task performance. Second, our results suggest that participants made order choices rapidly, at an early stage during the trials. Rapid utterance planning is generally thought to proceed via implicit planning decisions (Bock, 1996; Levelt, 1989). We propose that at least part of the bias we find emerged out of these implicit decisions rather than conscious strategizing.

However, one concern with our method is that the effects could stem from visual salience of the duplicated picture instead of, or in addition to, message uncertainty. We do not think that visual salience can entirely explain our results, given that earlier speech onset was found only for overlapping-first utterances in the Overlap condition. If visual salience were playing a major role in our experiment, we would expect to find that speakers are faster to begin speaking in the Overlap condition even when they name the nonoverlapping object first, because there is a duplicated picture in the display that can facilitate processing. Instead, we find that speech initiation latencies in the nonoverlapping first productions are equivalent to the Control trials, where there was no visual salience of a duplicated picture.

To further rule out the possibility of a visual salience confound, we designed Experiment 3. Instead of duplicated pictures, in Experiment 3 we use a semantically constraining question to provide early message information about the upcoming targets that participants are required to name. Using a semantically constraining question also creates a context that better resembles natural production contexts, improving ecological validity.

## Experiment 3

In Experiment 3, participants read a question under one of two conditions: 1) Semantic, where the question is semantically constraining and informative about the participant's upcoming response (e.g., *In the kitchen, what might be used to bake?*), and 2) Control, where the question does not provide any information about the upcoming targets (e.g., *In this display, what are the targets?*). Next, participants view four images on the screen. Of the four images in a given display, one image is strongly associated with the context in the Semantic question (e.g., oven), another two are plausible responses (e.g., apple, spoon), and one is unrelated (e.g., pill). After a brief preview, a cue indicates which two images are the targets to be named in a spoken utterance. In all trials, the two targets are the image that is strongly associated with the semantically constraining question (e.g., oven) and one of the other plausible responses (e.g., apple).

We hypothesize that in the Semantic condition participants will be more likely to name the strongly associated target first in their response, before the other target, producing utterances, such as *the oven and the apple*. This pattern would suggest that participants use the semantic information in the question to begin planning their responses in advance of the appearance of the visual cue indicating the targets. This result would suggest that the results of Experiment 1 and Experiment 2 were not solely due to the visual salience of the overlapping images, but rather reflect a strategy for language production under uncertainty. We also expect participants to begin speaking sooner in the Semantic condition compared to the Control condition, consistent with the results of the prior studies. Within the Semantic condition, we expect participants will begin speaking sooner when naming the strongly associated target first, again reflecting patterns in Experiments 1 and 2. This result would suggest that early utterance planning based on a semantic cue could provide a benefit in latencies, making it an efficient strategy for planning under uncertainty.

### Method

#### Participants

Undergraduate students (n = 101), all native-English speakers, participated in the experiment online for course credit. Sample size was determined by using a power analysis conducted using PANGEA (Westfall, 2016), aiming for at least 80% power with a medium effect size ($d$ = 0.5) in the analysis of word order. Data from 46 additional participants were excluded from analyses: 21 who had empty audio files, 14 who did not comply with task instructions, 7 who did not complete all trials, and 6 who had faulty audio files. Experiment 3 was preregistered before beginning data collection (https://osf.io/th9ck).

#### Stimuli

Object images from the MultiPic database (n = 160; Duñabeitia et al., 2017) were used as stimuli, assigned to 40 item displays of four objects each. Each display was assigned a Semantic question (e.g., *In the kitchen, what might be used to bake?*), providing a semantic cue for the upcoming target images, and a Control question, with no information about the upcoming images (e.g., *In this display, what are the targets?*). Of the four images in a given display, one image was strongly associated with the context in the Semantic question (e.g., oven), another two were plausible responses (e.g., apple, spoon), and one was unrelated (e.g., pill).

Before using the stimuli in our main experiment, we ran a norming study on a separate sample of 60 participants. We first created a list of the semantically constraining questions and four object images that could answer each question. Each participant was presented with each question and asked to rank the four images in order from 1 (the most appropriate for answering the question) to 4 (least appropriate). The image with the highest average ranking was then designated as the strongly associated object ($M$ = 1.42, $SD$ = 0.36). Images ranked third ($M$ = 2.82, $SD$ = 0.28) and fourth ($M$ = 3.42 , $SD$ = 0.34) were used as plausible responses. The image ranked second ($M$ = 2.21, $SD$ = 0.46) was discarded from that display in order to reduce competition from the first-ranked image and increase the strength of the manipulation. Instead, each second-ranked image was used as the distractor of another display, for a total of four images in each trial display: one strongly associated target, one plausible target, one plausible competitor (nontarget), and one distractor.

The mean number of words per question in the experiment was 7.42, $SD$ = 1.54 (Semantic $M$ = 8, $SD$ = 1.34; Control $M$ = 6.8, $SD$ = 1.54). Questions in the Semantic condition were designed to include the constraining information early in the first half of the question. Questions in the Control condition asked about images in the display without mentioning semantic content, such that they could be plausibly answered with any of the image names.

#### Procedure

The experiment was coded in jsPsych (de Leeuw, 2015) and run online via a lab server. Participants first read the task instructions on screen, including example displays and one practice trial with feedback, and then proceeded

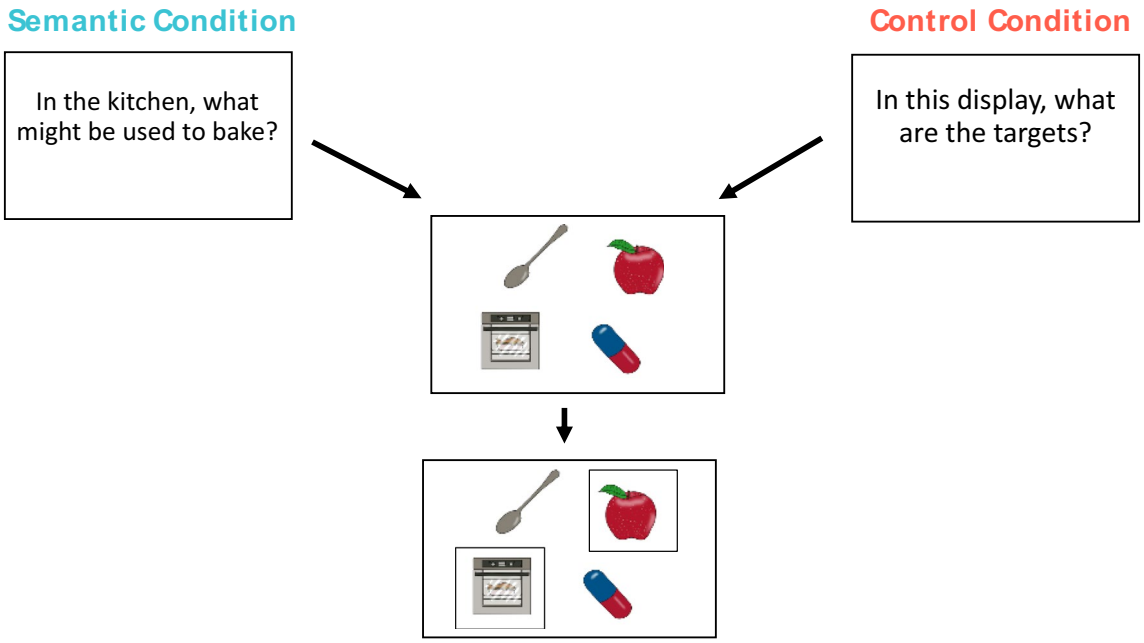


**Fig. 6** Example of a trial sequence under the Semantic and Control conditions. Participants first read the question on screen, and then were presented with the four images. The four images rotated clockwise to avoid screen position effects. After 3.8 seconds of a preview, square frames appeared around each of the target images, and participants named the targets in a conjoined noun phrase

to the experiment. Figure 6 illustrates the trial sequence. Each trial began with a question presented visually in the center of the screen, either in the Semantic or Control condition. Next, participants clicked on a green button in the center of the screen, and the question was replaced by a display of four images rotating clockwise in a circle. After a brief exposure of approximately 3.8 seconds, the two target images were framed with black squares. Participants' task was to name these two target images in a conjoined noun phrase ("the oven and the apple"), as a response to the prompt question. Participants were told that another participant would later listen to their recordings to identify the targets and that they should use the simple single-word labels for the objects (e.g., *oven* and not *gray oven*). The trial ended approximately 5.8 seconds after the targets were cued with the black frames.

Participants were randomly assigned to one of two lists, each consisting of twenty trials in the experimental condition and twenty trials in the control condition. Condition was counterbalanced across stimulus lists, such that each participant only viewed a given display once, but each display appeared in both conditions approximately equally across participants. The starting position of each image in each display was randomized by the experimental script. Trial order was randomized per participant. The entire experiment lasted less than 15 minutes.

## Analyses and results

### Preprocessing

Speech files were transcribed as in Experiment 1. Trials in which one or both objects were named inaccurately (38%) were removed from analyses. Montreal-Forced-Aligner (MFA; McAuliffe et al., 2017) was used to extract onset and offset times of each word in each utterance. At the time of the analysis for Experiment 3, MFA was more easily compatible with newer versions of Python compared with other aligners and therefore preferred. Before performing latency analyses, we further removed trials with disfluencies (6%), responses that did not follow the five-word conjoined noun phrase structure (65%), and trials where participants began speaking before the cue (8%). Note that percentages refer to the percentage of remaining observations after the prior step of data cleaning. This left 759 observations (372 Semantic, 387 Control; 19% of all observations) from 63 participants for each word position in the latency analyses.

### Word order

We first tested whether participants were more likely to name the strongly associated object first in the Semantic condition compared to the Control condition. We ran a mixed-effects logistic regression model regressing Order (strongly

associated object first vs. plausible object first) on Condition (Semantic vs. Control). The model included by-participant and by-item random intercepts and by-participant and by-item random slopes for Condition (the maximal random effects structure; Barr, 2013). As expected, participants were significantly more likely to produce the strongly associated object first in the Semantic condition compared with the Control (b = 0.27, *SE* = 0.1, $z = 2.75$, $p < 0.01$; Fig. 7).

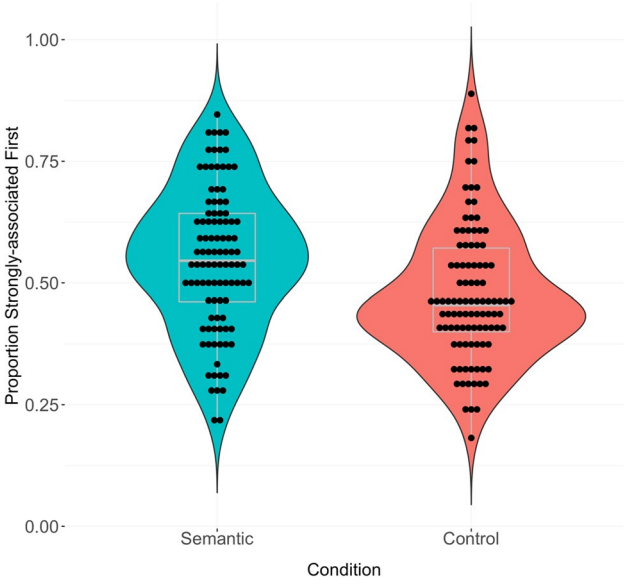


**Fig. 7** Violin plots of the proportion of trials in which participants named the strongly associated object first, in the Semantic condition and the Control condition in Experiment 3. Boxplots are overlayed in gray to identify quartiles and the median proportion. Points reflect individual participant means

## Latencies

We next examined speech initiation latencies (onsets) for each word in participants' utterances. As in the previous experiments, onsets were calculated as the time between the cue appearing and the first phoneme of each word and were log-transformed for analyses. The predictor variables were word Position (1-5), Condition (Control, Semantic; coded [0,1]), and Order of naming (plausible first, strongly associated first; coded [0,1]). Position was coded by using successive difference contrasts, where each word position is compared to the preceding word position (2-1, 3-2, 4-3, 5-4).

We ran a mixed-effects linear regression model regressing Onset on Position, Condition, Order, and their interaction. The random effects structure included by-subject and by-item random intercepts, and by-subject and by-item random slopes for Condition, Order, and their interaction. Significance values are reported using Satterthwaite's method *t*-tests.

There was no significant effect of Condition, Order, or their interaction (*p*s > 0.05; Fig. 8). Numerically, the estimated marginal means for onsets at the first word (latencies to begin speaking) were shorter when the strongly associated object (e.g., oven) was named first in the Semantic condition compared with all other cases. This pattern was maintained throughout all five-word positions and even increased with sentence progression. The numerical pattern therefore aligns

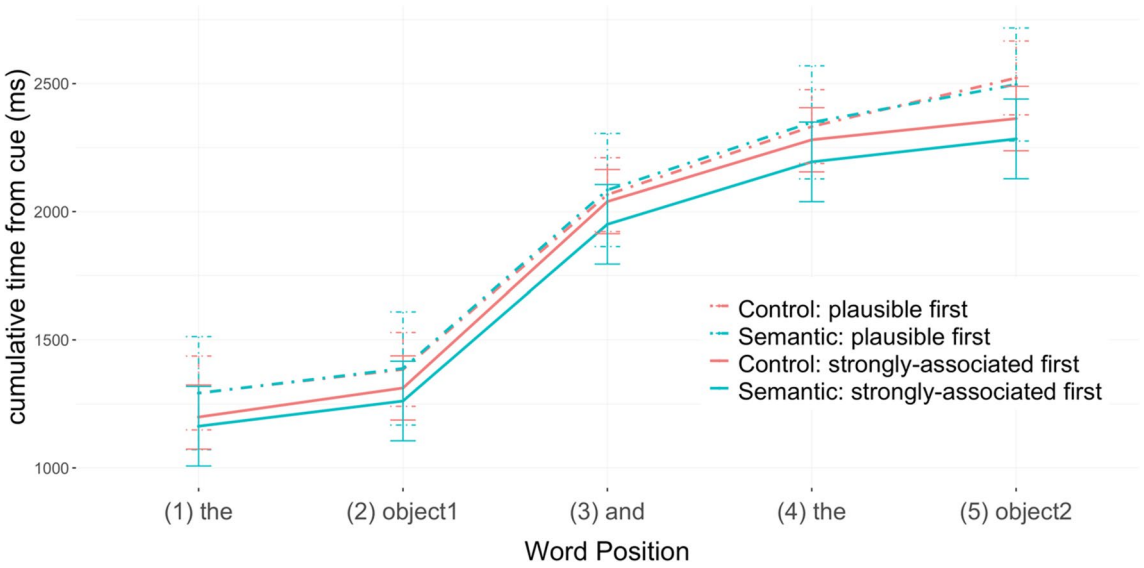


**Fig. 8** Model predictions for onsets in Exp 3. Data are divided into trials where participants named the strongly associated target first (e.g., oven; solid lines) or the plausible target first (e.g., apple; dashed lines), in the Semantic condition (blue) or the Control condition (pink). Error bars represent ±1 SE

with our hypothesis, but the interaction between Order and Condition was not reliable.

### Experiment 3 discussion

Results from Experiment 3 align with our findings in Experiments 1 and 2. Participants were more likely to name the strongly associated object first in the Semantic condition, suggesting they used information from the preceding question to begin planning their responses early—affecting word order choices. This finding supports our argument that visual salience alone cannot explain our results in Experiments 1 and 2, as there were no duplicated images in Experiment 3 that could affect participants' behavior.

Also note that in Experiments 1 and 2, the Overlap trials allowed participants to predict with complete accuracy what one of the targets was going to be (the overlapping target), while uncertainty remained around the second target. In Experiment 3, participants did not have complete certainty about any of the targets but appear to have used semantic-based predictions to begin planning the most likely target first (the strongly associated target). In both cases, we find that speakers plan their utterances using any available information, indicating a robust effect of message uncertainty on utterance formulation, across these variations in uncertainty type. Moreover, the results of Experiment 3 suggest that decisions about word order and early planning do not necessarily require complete certainty about a specific message component for utterance planning to begin but rather could be based on speakers' predictions. Presumably, a strong enough prediction is worth the risk of an error in utterance planning (i.e., in case a planned word is not the correct target). Although we were underpowered to analyze changes in productions over multiple trials, future studies could examine how participants' strategies evolve over the course of the experiment, e.g., whether they rely more strongly on their predictions for early planning, learn to refine their predictions, and more.

While the word order effect was robust across experiments, results from the latency analysis in Experiment 3 were inconclusive: the numerical pattern aligned with our hypothesis—shortest speech initiation latencies were found for trials in the Semantic condition where the strongly associated object was named first, but the difference was not statistically significant. It is possible that we had too few observations to detect the expected interaction effect (between Condition and Order) after data exclusions. Due to constraints for finding item sets to align with the Semantic question condition, Experiment 3 had fewer items than in Experiments 1 and 2. Moreover, data exclusions were more common due to varying sentence structures and inaccurate object naming. The higher exclusion rate is likely because of the online format of our experiment, which had no familiarization phase for the object names and no interaction with research personnel. All of these factors could decrease the motivation to respond quickly and accurately, and decrease adherence to experiment instructions more generally. Future experiments might maximize the manipulation strength by using a confederate interlocutor or a deadline that encourages participants to begin speaking more quickly or to exert more control on participants' sentence structures to maximize power for detecting an effect. Nonetheless, the results of the word-order analysis were robust, supporting our hypothesis that participants use early semantic information to begin planning their responses even when only partial message information is available.

## General discussion

We developed two novel methods to manipulate the uncertainty of producers' messages in a picture-naming task. Results showed that word order choices varied with degree of message uncertainty: producers prioritized message components likely to be needed, resulting in early placement of these elements in the utterance. These findings suggest that speakers make early utterance decisions based on available information, allowing them to begin speaking sooner, and then continue planning the rest of the utterance when more information becomes available. Our results were comparable across spoken and typed modalities, suggesting similar strategies for planning under uncertainty, regardless of output modality.

Although this initial investigation was necessarily more constrained than natural conversations, it is likely that similar patterns will emerge in conversation. Speakers often begin planning their response to an interlocutor before the interlocutor has finished speaking (Bögels, 2020; Bögels et al., 2015; Corps et al., 2018), suggesting that sometimes utterance planning precedes knowledge of the full information needed to reply. To explain how speakers begin early planning in conversation, turn-taking researchers typically point to predictive comprehension, which suggests that comprehenders can use context cues to predict what the upcoming words or messages are going to be (Kuperberg and Jaeger, 2016). Using these predictions, speakers can begin planning their own response even before the interlocutor has finished their turn (Corps et al., 2018; Levinson, 2016; Levinson and Torreira, 2015).

But predictions are, by definition, uncertain. Depending on the degree of constraint in the sentence, the degree of the listener's confidence in their prediction may vary (Klimovich-Gray et al., 2019; Lewis and Bastiaansen, 2015; Luke and Christianson, 2016). Predictions also can be somewhat vague, i.e., the listener might have a general idea of what the speaker is going to say but still be unsure how exactly the

utterance will turn out. Even trained coders often disagree on when exactly in an incoming question the answer becomes clear (Bögels, 2020). This result suggests that response planning based on predictions also must carry some uncertainty; if the speaker does not know exactly what they are responding to, they cannot know exactly what to plan in their response utterance, and production strategies might vary (Gussow, 2023).

Our study expands on these ideas in several ways. First, prior turn-taking work focused on stimuli with high message predictability, e.g., general knowledge questions with short answers, manipulating whether the required answer became clear mid-question (allowing early planning) or only at the end (Bögels et al., 2015). In Experiments 1 and 2 presented, timing was carefully controlled, while we manipulated *how much* information participants had about their message, and consequently, *how much* of their utterance they could plan. This type of message uncertainty resembles conversational contexts where speakers begin planning their response when they only have partial information about what they are responding to. In Experiment 3, we introduced early semantic cues indicating *how likely* certain components are to be included in the message and found that speakers prioritize those likely components in their response. Using semantically constraining questions in Experiment 3 allowed a more ecologically valid context for our experiment, as natural conversation similarly contains a semantic context that constrains speakers' messages. Notably, although our participants were told to respond promptly, they had plenty of time to produce their utterances and no penalty for slow responses. The fact that they still attempted early planning, even under uncertainty, might suggest that this production strategy is well-practiced from their own language experience, becoming natural even without obvious time pressure.

Another novel aspect of our findings is that uncertainty affected participants' utterance forms, as they chose word orders that allowed early planning, potentially over alternatives that would have prioritized the listener's needs. Recall that participants were told that another participant would later listen to their recordings to identify the targets. From an audience-design perspective emphasizing getting the information to the listener as early as possible, naming the *nonoverlapping* object first in Experiments 1 and 2 would be the most efficient word order for the listener to hear, because a nonoverlapping object immediately identifies the target pair and allows the listener to complete the task goal. Similarly, naming the *less likely* object in Experiment 3 (the plausible object, but not the strongly associated object) would be more informative to the listener, because the listener could be quite confident that the strongly associated object would be a target, but more uncertain about which of the plausible objects would be the other target. Participants still named the overlapping (or strongly associated) object first, suggesting that early planning for the speaker was prioritized over early informativity for a future listener, consistent with studies suggesting that there are limits to the degree to which producers accommodate listeners' needs (Horton and Keysar, 1996). Indeed, utterance form choices often reflect speakers' use the flexibilities of language to mitigate production demands (MacDonald, 2013), and planning under uncertainty is a particular type of production demand.

Given the current results, natural further steps include testing participants in conversational settings with interlocutors for a more ecologically valid context, and varying the degree or type of message uncertainty to see how planning biases are affected. Another step is to begin exploring the neural correlates of production under message uncertainty, especially given that effects of competition in early planning can sometimes be detected in the neural signal but not in behavioral measures, such as reaction time (Piai et al., 2020). More generally, more research on production under uncertainty would not only address a common everyday situation that is understudied in the laboratory but also could shed light on the interaction between various production processes, including early planning, message formulation, and word-order choices.

Notably, strategic yet implicit decision-making in the face of uncertainty is not unique to language production and our work furthers relationships between decision making in language and other areas. In fact, motivation for our study drew on other decision-making contexts (Chapman et al., 2010; Gallivan et al., 2015), where producers adapt their motor action plans to cope with uncertainty about alternative target options. There also is a rich body of work on the neural correlates of goal uncertainty during action planning, showing how neuronal activity is tuned to particular goals and modulated by goal uncertainty. For example, when monkeys are provided with only partial information about upcoming reach options, activity in reach-related neural populations is modulated by the degree of initial uncertainty (determined by the number of competing targets), and changes dynamically as the response time approaches (Bastian et al., 2003). In human participants, target uncertainty modulates oscillatory activity during motor response selection (van Helvert et al., 2021), and modulation is based on the number of potential targets and their degree of similarity (Grent et al., 2015; Tzagarakis et al., 2010, 2015).

Similarly, in the language domain, modulation of neural activity has been found when there is competition between multiple production plans (Marian et al., 2017; Piai et al., 2014), although to our knowledge work of this sort has not yet addressed message uncertainty. Also EEG data suggest that in comprehension, multiple word predictions can be represented in a graded matter depending on their likelihood (DeLong et al., 2005; but see Nieuwland et al., 2018), indicating sensitivity to the

degree of uncertainty in comprehension. The evidence is limited however, and there is a need for research on *message* competition in production, as with goal competition in the motor domain. We suggest that investigations of language production could benefit greatly not only from examining how traditional language-related brain networks (Friederici, 2011; Indefrey, 2011) or electrophysiological markers of early planning (Bögels et al., 2015; Piai and Zheng, 2019) are modulated based on the degree or type of message uncertainty, but also from parallels in motor action and other decision making areas (Gussow, 2023). The novel yet simple paradigms we have presented would be a useful tool to begin these investigations, because they can likely be adapted to studies with physiological measures.

Both language production and motor action are often viewed as continuous decision making, and several researchers have previously pointed to similarities between the domains (Anderson and Dell, 2018; Beaty et al., 2020; Gussow, 2023; Koranda et al., 2020; Lashley, 1951; Lebkuecher et al., 2022; MacDonald, 2016). Although different cognitive domains have their own sets of task demands, parallels in higher-level planning strategies do exist and are mutually informative—helping to constrain the hypothesis space and lend ideas for experimental manipulations of uncertainty in other domains. Together, the research follows an approach that not only benefits each domain individually but also could increase our understanding of domain-general neurocognitive processes, in particular in the context of message and goal uncertainty.

**Acknowledgments** This research was supported by National Science Foundation Grant 1849326. The authors thank research assistants in the Language and Cognitive Neuroscience Lab for help with data collection, transcription, and coding.

**Funding** This research was supported by National Science Foundation Grant 1849326.

**Data availability** Data and trial demonstrations are available online at https://osf.io/zu9m8/.

**Code availability** The R script for statistical analyses is available online at https://osf.io/zu9m8/.

## Declarations

**Ethics approval** The experiments were reviewed and approved by Social Sciences IRB, University of Wisconsin-Madison.

**Consent to participate** Informed consent was obtained from all individual participants included in the study.

**Consent for publication** As part of the informed consent procedure, participants were made aware that their deidentified data would be submitted for publication.

**Conflicts of interest** The authors have no conflicts of interest to declare.

## References

Allum, P. H., & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: The role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 791–810. https://doi.org/10.1037/0278-7393.33.4.791

Anders, R., Riès, S., van Maanen, L., & Alario, F.-X. (2015). Evidence accumulation as a model for lexical selection. *Cognitive Psychology, 82*, 57–73. https://doi.org/10.1016/j.cogpsych.2015.07.002

Anderson, N. D., & Dell, G. S. (2018). The role of consolidation in learning context-dependent phonotactic patterns in speech and digital sequence production. *Proceedings of the National Academy of Sciences, 115*(14), 3617–3622. https://doi.org/10.1073/pnas.1721107115

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research, 32*(1), 25–36.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445–459. https://doi.org/10.3758/BF03193014

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00328

Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01858

Bastian, A., Schöner, G., & Riehle, A. (2003). Preshaping and continuous evolution of motor cortical representations during movement preparation. *European Journal of Neuroscience, 18*(7), 2047–2058. https://doi.org/10.1046/j.1460-9568.2003.02906.x

Battaglia, P. W., & Schrater, P. R. (2007). Humans trade off viewing time and movement duration to improve visuomotor accuracy in a fast reaching task. *Journal of Neuroscience, 27*(26), 6984–6994.

Beaty, R., Frieler, K., Norgaard, M., Merseal, H., MacDonald, M., & Weiss, D. (2020). *Spontaneous melodic productions of expert musicians contain sequencing biases seen in language production*. https://doi.org/10.31234/osf.io/qdh32

Bock, K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*(1), 1–47. https://doi.org/10.1037/0033-295X.89.1.1

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*(3), 355–387.

Bock, K. (1995). Sentence production: From mind to mouth. In *Speech, language, and communication* (pp. 181–216). Academic Press. https://doi.org/10.1016/B978-012497770-9/50008-X

Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review, 3*(4), 395–421. https://doi.org/10.3758/BF03214545

Bögels, S. (2020). Neural correlates of turn-taking in the wild: Response planning starts early in free interviews. *Cognition, 203*, 104347. https://doi.org/10.1016/j.cognition.2020.104347

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports, 5*(1), 12881. https://doi.org/10.1038/srep12881

Bögels, S., Casillas, M., & Levinson, S. C. (2018). Planning versus comprehension in turn-taking: Fast responders show reduced anticipatory processing of the question. *Neuropsychologia, 109*, 295–310. https://doi.org/10.1016/j.neuropsychologia.2017.12.028

Branigan, H. P., & McLean, J. F. (2016). What children learn from adults' utterances: An ephemeral lexical boost and persistent syntactic priming in adult–child dialogue. *Journal of Memory and Language, 91*, 141–157. https://doi.org/10.1016/j.jml.2016.02.002

Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition, 109*(2), 274–280. https://doi.org/10.1016/j.cognition.2008.07.011

Brown-Schmidt, S., & Konopka, A. E. (2015). Processes of incremental message planning during conversation. *Psychonomic Bulletin & Review, 22*(3), 833–843. https://doi.org/10.3758/s13423-014-0714-2

Chapman, C. S., Gallivan, J. P., Wood, D. K., Milne, J. L., Culham, J. C., & Goodale, M. A. (2010). Reaching for the unknown: Multiple target encoding and real-time decision-making in a rapid reach task. *Cognition, 116*(2), 168–176. https://doi.org/10.1016/j.cognition.2010.04.008

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

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. https://doi.org/10.1080/0163853X.2017.1330031

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117–1121. https://doi.org/10.1038/nn1504

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2017). MultiPic: A standardized set of 750 drawings with norms for six European languages. *The Quarterly Journal of Experimental Psychology, 0*(ja), 1–24. https://doi.org/10.1080/17470218.2017.1310261

Faisal, A. A., & Wolpert, D. M. (2009). Near optimal combination of sensory and motor uncertainty in time during a naturalistic perception-action task. *Journal of Neurophysiology, 101*(4), 1901–1912. https://doi.org/10.1152/jn.90974.2008

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language, 46*(1), 57–84. https://doi.org/10.1006/jmla.2001.2797

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews, 91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Gallivan, J. P., Barton, K. S., Chapman, C. S., Wolpert, D. M., & Randall Flanagan, J. (2015). Action plan co-optimization reveals the parallel encoding of competing reach movements. *Nature Communications, 6*(1), 7428. https://doi.org/10.1038/ncomms8428

Garrett, M. F. (1989). Processes in language production. *Linguistics: The Cambridge Survey, 3*, 69–96.

Gennari, S. P., Mirković, J., & MacDonald, M. C. (2012). Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive Psychology, 65*(2), 141–176. https://doi.org/10.1016/j.cogpsych.2012.03.002

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language, 57(*4), 544–569. https://doi.org/10.1016/j.jml.2007.01.007

Grent, T., Oostenveld, R., Medendorp, W. P., & Praamstra, P. (2015). Separating visual and motor components of motor cortex activation for multiple reach targets: A visuomotor adaptation study. *Journal of Neuroscience, 35*(45), 15135–15144.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition, 82*(1), B1–B14. https://doi.org/10.1016/S0010-0277(01)00138-X

Gussow, A. E. (2023). Language production under message uncertainty: When, how, and why we speak before we think. In K. D. Federmeier & J. L. Montag (Eds.), *Psychology of learning and motivation: Speaking, writing, and communicating (Vol. 78)*. Academic Press. https://doi.org/10.1016/bs.plm.2023.02.005

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science, 8*(3), 191–219. https://doi.org/10.1016/S0364-0213(84)80001-4

Heller, D., Skovbroten, K., & Tanenhaus, M. K. (2009). Experimental evidence for speakers sensitivity to common vs. privileged ground in the production of names. *Proc PRE-CogSci.*

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*(1), 91–117. https://doi.org/10.1016/0010-0277(96)81418-1

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology, 2*, 255. https://doi.org/10.3389/fpsyg.2011.00255

Jongman, S. R., Piai, V., & Meyer, A. S. (2020). Planning for language production: The electrophysiological signature of attention to the cue to speak. *Language, Cognition and Neuroscience, 35*(7), 915–932. https://doi.org/10.1080/23273798.2019.1690153

Klimovich-Gray, A., Tyler, L. K., Randall, B., Kocagoncu, E., Devereux, B., & Marslen-Wilson, W. D. (2019). Balancing prediction and sensory input in speech comprehension: The spatiotemporal dynamics of word recognition in context. *The Journal of Neuroscience, 39*(3), 519–527. https://doi.org/10.1523/JNEUROSCI.3573-17.2018

Koranda, M. J., Bulgarelli, F., Weiss, D. J., & MacDonald, M. C. (2020). Is language production planning emergent from action planning? *A Preliminary Investigation. Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.01193

Koranda, M. J., Zettersten, M., & MacDonald, M. C. (2022). Good-enough production: Selecting easier words instead of more accurate ones. *Psychological Science*, *09567976221089603*. https://doi.org/10.1177/09567976221089603.

Krüger, M., & Hermsdörfer, J. (2019). Target uncertainty during motor decision-making: The time course of movement variability reveals the effect of different sources of uncertainty on the control of reaching movements. *Frontiers in Psychology, 10*, 41. https://doi.org/10.3389/fpsyg.2019.00041

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–131). Wiley.

Lebkuecher, A. L., Schwob, N., Kabasa, M., Gussow, A. E., MacDonald, M. C., & Weiss, D. J. (2022). Hysteresis in motor and language production. *Quarterly Journal of Experimental Psychology, 17470218221094568*. https://doi.org/10.1177/17470218221094568

Levelt, W. (1989). Speaking: From intention to articulation MIT Press. Cambridge, MA.

Levinson, S. C. (2016). Turn-taking in human communication – Origins and implications for language processing. *Trends in Cognitive Sciences, 20*(1), 6–14. https://doi.org/10.1016/j.tics.2015.10.010

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology, 6*, 731.

Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex, 68*, 155–168. https://doi.org/10.1016/j.cortex.2015.02.014

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology, 88*, 22–60. https://doi.org/10.1016/j.cogpsych.2016.06.002

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00226

MacDonald, M. C. (2016). Speak, act, remember: The language-production basis of serial order and maintenance in verbal memory. *Current Directions in Psychological Science, 25*(1), 47–53. https://doi.org/10.1177/0963721415620776

Marian, V., Bartolotti, J., Rochanavibhata, S., Bradley, K., & Hernandez, A. E. (2017). Bilingual cortical control of between-and within-language competition. *Scientific Reports, 7*(1), 11763.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using Kaldi. *Interspeech, 2017*, 498–502.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., & Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife, 7*, e33468

Piai, V., & Zheng, X. (2019). Chapter fight - speaking waves: Neuronal oscillations in language production. In K. D. Federmeier (Ed.), *Psychology of learning and motivation* (Vol. 71, pp. 265–302). Academic Press. https://doi.org/10.1016/bs.plm.2019.07.002

Piai, V., Roelofs, A., Jensen, O., Schoffelen, J.-M., & Bonnefond, M. (2014). Distinct patterns of brain activity characterise lexical activation and competition in spoken word production. *PLOS ONE, 9*(2), e88674

Piai, V., Klaus, J., & Rossetto, E. (2020). The lexical nature of alpha-beta oscillations in context-driven word production. *Journal of Neurolinguistics, 55*, 100905. https://doi.org/10.1016/j.jneuroling.2020.100905

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*(3), 460–499. https://doi.org/10.1037/0033-295X.107.3.460

Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 353–368. https://doi.org/10.1037/0278-7393.34.2.353

Rosenbaum, D. A., Cohen, R. G., Jax, S. A., Weiss, D. J., & van der Wel, R. (2007). The problem of serial order in behavior: Lashley's legacy. *Human Movement Science, 26*(4), 525–554. https://doi.org/10.1016/j.humov.2007.04.001

Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., & Yuan, J. (2014). FAVE (forced alignment and vowel extraction). *Program Suite, v1*, 2.2.

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. *Analyzing discourse: Text and talk, 71*, 71–93.

Smith, M., & Wheeldon, L. (2004). Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(3), 675–686. https://doi.org/10.1037/0278-7393.30.3.675

Snyder, K. M., & Logan, G. D. (2014). The problem of serial order in skilled typing. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1697–1717. https://doi.org/10.1037/a0037199

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

Tzagarakis, C., Ince, N. F., Leuthold, A. C., & Pellizzer, G. (2010). Beta-band activity during motor planning reflects response uncertainty. *Journal of Neuroscience, 30*(34), 11270–11277.

Tzagarakis, C., West, S., & Pellizzer, G. (2015). Brain oscillatory activity during motor preparation: Effect of directional uncertainty on beta, but not alpha, frequency band. *Frontiers in Neuroscience, 9*, 246.

van Helvert, M. J. L., Oostwoud Wijdenes, L., Geerligs, L., & Medendorp, W. P. (2021). Cortical beta-band power modulates with uncertainty in effector selection during motor planning. *Journal of Neurophysiology, 126*(6), 1891–1902. https://doi.org/10.1152/jn.00198.2021

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin, 128*, 442–472.

Westfall, J. (2016). *PANGEA: Power ANalysis for GEneral Anova designs (working paper)*.

Wolpert, D. M., & Landy, M. S. (2012). Motor control is decision-making. *Current Opinion in Neurobiology, 22*(6), 996–1003. https://doi.org/10.1016/j.conb.2012.05.003

Wong, A. L., & Haith, A. M. (2017). Motor planning flexibly optimizes performance under uncertainty about task goals. *Nature Communications, 8*(1), 1. https://doi.org/10.1038/ncomms14624

**Open practices statements** Data are available online at https://osf.io/zu9m8/. Experiments 1 and 2 were not preregistered; Experiment 3 was preregistered at https://osf.io/th9ck before data collection.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.