Investigating fragment usage with a gamified utterance selection task

Robin Lemke*

Abstract. Nonsentential utterances, or fragments, like *A coffee, please!* can often be used to communicate a propositional meaning otherwise encoded by a complete sentence (*I'd like to order a coffee, please!*). Previous research focused mostly on the syntax and licensing of fragments, but the questions of why speakers use fragments and how listeners interpret them are still underexplored. I propose a simple game-theoretic account of fragment usage, which predicts (i) that listeners assign fragments the most likely interpretation in context and (ii) that speakers are aware of this and trade-off production cost and the risk of being misunderstood when choosing their utterance. Using a corpus of production data, empirically founded and precise model predictions are generated. These predictions are evaluated with two experiments using a novel gamified utterance selection paradigm. The experiments suggest that, as predicted, speakers take into account both potential gains in efficiency and the risk of being misunderstood when choosing their utterance.

Keywords. ellipsis, fragments, game theory

1. Introduction. Instead of a complete sentence like (1a), speakers often use nonsentential utterances, or fragments (Morgan 1973), for this purpose (1b). Despite their reduced form, fragments can be meaning-equivalent to their fully sentential counterparts in an appropriate context.

(1) [passenger asks conductor on the platform next to a waiting train:]
   a. Does this train go to Paris?
   b. To Paris?

While the syntax of fragments has been and is being extensively studied (see e.g. Ginzburg & Sag 2000, Merchant 2004, Reich 2007, Weir 2014, Ott & Struckmeier 2016, Lemke 2021), the question of why speakers actually choose to use a fragment is relatively underexplored (but cf. Bergen & Goodman 2015, Lemke 2021).

Intuitively, fragments have the advantage of being shorter than sentences, which reduces the production effort for the speaker and the (syntactic) processing effort for the listener. This makes communication more efficient, as the same information is transmitted in less time, a tendency that also can be attributed to the Gricean maxim of manner "be brief" (Grice 1975). The downside of fragments is their vagueness: A single fragment can often be used to communicate meanings expressed by different complete sentences. For instance, the fragment in (1b) can communicate not only (1a), but also any of the questions in (2). Due to this ambiguity, the listener might assign the fragment a different interpretation than intended by the speaker, or they might not be able to retrieve any interpretation at all. Both of these outcomes would result in communication failure and require further clarifications, making communication less efficient.

(2) a. Are you traveling to Paris?
   b. Have you ever been to Paris?

*Authors: Robin Lemke, Saarland University (robin.lemke@uni-saarland.de).

I pursue the hypothesis that speakers counterbalance the gain in efficiency provided by fragments with the risk of communication failure caused by their ambiguity when choosing how to encode the message. If the gain in efficiency outweighs the risk, the speaker will use the fragment; if it does not, a full sentence. This might explain the observation in Lemke (2021) that fragments are preferred over sentences as answers to questions, but the opposite holds in discourse-initial contexts, where the uncertainty about their intended interpretation is probably greater.

To formalize this idea, I use a simple game-theoretic model, which allows for including an explicit utterance cost structure and which predicts how likely it actually is to communicate a particular meaning in a context. Such models, which have been widely applied to pragmatic phenomena, such as scalar implicature (Franke 2009), reference (Frank & Goodman 2012, Rohde et al. 2012) or the construction of social meaning (Burnett 2017), predict speaker and listener behavior based on an explicit model of the utterance context. From the perspective of ellipsis research, this has the advantage of explaining the interpretation of elliptical utterances based on pragmatic inferences (i.e. relying on the same mechanisms as implicature generation), instead of having to assume a specific processing mechanism, as has been suggested by e.g. Arregui et al. (2006) for VP ellipsis. In order to generate model predictions, I rely on a corpus of production data collected by Lemke (2021). This approach differs from most previous empirical investigations of game-theoretic reasoning in pragmatics, which relied on controlled and balanced experimental setups in which objects are referred to by their shape or color with one- or two-word utterances (e.g. Frank & Goodman 2012, Rohde et al. 2012, Sikos et al. 2021). To my knowledge, my study is therefore the first attempt in game-theoretic pragmatics to derive model predictions from a relatively large, unbalanced, and diverse data set based on linguistic data produced by human subjects.

This paper is organized as follows: Section 2 presents the game-theoretic account of fragment usage and Section 3 the data set on which model predictions are based. Section 4 introduces the experimental design, before Sections 5 and 6 present the results of the two experiments and Section 7 summarizes the main results and conclusions.

2. A game-theoretic account of fragment usage. I formalize the idea that fragment usage results from a trade-off between the gain in efficiency and the risk of being misunderstood within a signaling game framework (Franke 2009, Frank & Goodman 2012, Sikos et al. 2021). The model predicts speaker and listener behavior based on a (possibly partial) mapping between a set of utterances and a set of messages, i.e. meanings the speaker could communicate. The components of the model are mostly equivalent to those in Franke (2009), although I slightly adapt the terminology to better fit the inference from (possibly nonsentential) utterances to complete sentences.

There is a set of messages $M \in m$ comprising all of the meanings the speaker can communicate in a situation (e.g. (1a) and (2) in the train example), and a set of utterances $u \in U$ that they can use for this purpose. The speaker's task is to send the utterance which is optimal to communicate the intended message. The listener receives this utterance and performs an interpretation action $a \in A$, i.e. they assign the utterance a meaning. Since M contains all possible meanings, $A = M$. The set of utterances U contains (i) sentences, which are equivalent to the $m \in M$[1] and

[1]This is a simplification, because there are also meaning-equivalent sentences. I address this concern by pooling the production data to sentences differing in meaning as discussed in Section 3, but it would also be possible to include all of the synonym sentences in U.

(ii) all fragments which can be derived from the $m \in M$ by grammatically licensed omission.[2] I consider only fragments that are constituents or sequences thereof (e.g. *To Paris?*, *This (one) to Paris?*, but not, for example, the heads of a DP and a PP only (e.g. *This to?*). The reason for this is that fragments are licensed by focus or givenness (e.g. Merchant 2004, Reich 2007) and these concepts apply generally to constituents. Having established M and U, it is necessary to determine whether an utterance u can be used to communicate a message m. In his account of implicature, Franke (2009) relies on truth conditions, but fragments cannot be considered true or false before having been enriched to a complete proposition. Given the computation of U described above, I rely on whether u can be derived from m by grammatically licensed omission. Following the notation in Franke (2009), assume a denotation function $[[\cdot]]$, which returns 1 if u can be derived from m and 0 otherwise. Finally, not all messages are necessarily equally likely in a situation. This might be due to world knowledge (people are more likely to ask the conductor whether the train goes to Paris than whether they have been to Paris) or statistical knowledge about language use. This potential difference in likelihood is captured by a prior probability distribution $Pr(M)$.

Based on these components, following Franke (2009), chains of mutual reasoning about the behavior, preferences and beliefs of the interlocutors are initialized: one starting from a "literal" speaker S_0 and one starting from a "literal" listener L_0. Since I model the speaker's perspective in my experiments, I focus on the basic listener model, which the speaker should consider when selecting their utterance. Given a message u sent by the speaker, L_0 calculates the likelihood of each $m \in M$ given u. As equation 1 shows, this posterior $p(m|u)$ is determined as the ratio of $Pr(m)$ and the probability mass of other possible interpretations m' of u. This reweighs the prior probability among those messages from which u could have been derived. Based on the resulting probability distribution, the listener selects the most likely message as the interpretation of u. As I discuss in Section 4.2, the reason for this is that successful communication is rewarded with a payoff and maximizing $L_0 p(m, u)$ also maximizes the payoff.

$$L_0(m, u) = \frac{Pr(m) \times [[u]]_m}{\sum_{m'} Pr(m') \times [[u]]_{m'}} \tag{1}$$

A speaker who takes listener behavior into account (a S_1 speaker in Franke's terminology) will constrain their utterance choice based on the likelihood that the listener will interpret it as intended. Since U has been derived by omission from M, sentential utterances fully disambiguate between messages. Therefore, from the perspective of successful communication they should be always at least as ideal as a fragment. However, sentences come with an additional production cost, which might result in them being less ideal for the speaker.[3] Since it is a priori unclear how to quantify utterance cost and its relationship to the gain in efficiency, I use an explicit cost structure in my experiments to ensure that sending fragments is cheaper than sending sentences.

[2]Note that I follow most of the theoretical literature cited above and treat fragments as grammatically well-formed, but there are diverging views: Bergen & Goodman (2015) assume that fragments are ungrammatical but can still be suitable means of communication if the listener can retrieve their meaning. The account I propose is in principle independent from the question of whether fragments are grammatical, but assuming ungrammaticality would require to include ungrammatical fragments in U, which I do not.

[3]Of course, unnecessary redundancy might be dispreferred by the listener (Schäfer et al. 2021), but since they have no control over which utterance the speaker sends, this does not need to be modeled.

3. Data set. Testing the predictions of the account described in Section 2 empirically requires realistic estimates of the model's components: (i) the set of messages M, (ii) the set of utterances U, (iii) the mapping between M and U, and (iv) the prior probability distribution over messages $Pr(M)$. I calculate these parameters from a data set originally collected by Lemke (2021).[4] Lemke (2021) used a crowd-sourced written production task to elicit utterances with script-based context stories like (3). The participants were instructed to produce the most natural utterance in that situation by the person specified in the prompt *Annika to Jenny*:

(3) Annika and Jenny want to cook pasta. Annika has put a pot with water on the stove. Then she has turned the stove on. After a few minutes, the water has started to boil. Annika to Jenny:

The stories were based on event chains extracted from the DeScript corpus of script knowledge (Wanzare et al. 2016). For each story, which represents a different script-based scenario, about 100 responses were collected. The answers (e.g. (4a)) were preprocessed into abstract representations like (4b), where each word corresponded to a constituent that could be freely omitted in German. This ensures that all fragments derived from these representations are grammatical, i.e. constituents or sequences thereof, as discussed in Section 2.[5]

(4) a. Pour the pasta into the pot!
b. `pour pasta pot.GOAL`

The set of unique representations for each scenario (e.g. *cooking pasta*) was taken to represent M in this scenario. Since German has free word order, all of the words in the representations like (4b) were ordered alphabetically. This ensures that the synonym `pour pasta pot.GOAL` and `pour pot.GOAL pasta` are not treated as different meaning representations. The likelihood of each representation within the 100 preprocessed responses determines $Pr(M)$ in this scenario. U is the union of M and all fragments that can be derived by grammatical omission from all $m \in M$: For a message m, this includes the corresponding sentence, all of its constituents (i.e. the individual "words" in the abstract representation) and all possible combinations of thereof. (5) exemplifies this at the example of (4b).

(5) `pour pasta pot.GOAL`, `pasta pot.GOAL`, `pour pot.GOAL`, `pour pasta`, `pour`, `pasta`, `pot.GOAL`

Finally, $[[u]]_m$ determined for each $u \in U$ and each $m \in M$ whether u could be derived from m by omission. Applying equation 1, the likelihood of communicative success $L_0(m|u)$ was then calculated for each $m \in M$ and $u \in U$.

4. Experimental approach. The experiments presented below were designed to test the hypothesis that the usage of fragment is conditioned by a trade-off between the gain in efficiency and the risk of being misunderstood. Therefore, fragments are expected to be more strongly preferred if (i) their cost relative to a sentence is lower, and if (ii) the likelihood of communicative success is relatively high.

[4]The data set was collected in German, but I present the context stories in English for convenience.
[5]For details on the preprocessing procedure, see Lemke (2021; 202–206).

I tested this with a pseudo-interactive gamified utterance selection task inspired by Rohde et al. (2012). In my experiments, the participant took the speaker role and chose between different sentential and fragment utterances to communicate a message determined by the experiment. The listener was simulated to behave according to model predictions: For sentences, they always selected the correct interpretation and for fragments, they maximized $L_0p(m|u)$. An explicit cost structure ensured that fragments were cheaper than sentences.

4.1. MATERIALS: CONTEXT STORY, MESSAGES, AND UTTERANCES. The stimuli were derived from the data set by Lemke (2021) introduced above. In each trial, the participant saw a context story, three messages and six utterances, as shown in Figure 1. The context stories were based on those used by Lemke (2021), but the roles of the characters were assigned to the participant and their (simulated) partner. The messages were presented as states of affairs that the participant might want to communicate to their partner. The message the participant had to communicate was determined by the experiment depending on the experimental condition (See Section 4.2) and highlighted in blue. Among the six utterances, three were sentences, each unambiguously encoding one of the messages. One of the three fragments (*into the water* in Figure 1) is ambiguous and can communicate two of the messages. These two messages were selected from the production data so that one of them has a higher $L_0p(m|u)$ given the ambiguous fragment than the other one. If participants base their utterance choice on the likelihood of communicative success, they should be more likely to use a fragment when asked to refer to the more likely message than when referring to the less likely one, because a listener maximizing $L_0p(m|u)$ will choose the more likely interpretation more often. The second fragment (in the example: *on the table*) always referred unambiguously to the third message. This establishes a baseline for the rate of fragment usage when there is no risk of misunderstanding in the experimental setup. The third fragment (*the recipe!*) had no corresponding message and served as a control to exclude inattentive participants who randomly click on any of the utterances. All messages and utterances were based on the abstract representations contained in the production data for the respective scenario.

4.2. CONDITIONS, UTTERANCE COST AND EXPECTATIONS. There were three conditions differing in the message to be communicated. In the *critical* condition, the most likely message given the ambiguous fragment was highlighted. In the *competitor* condition, the less likely message compatible with the ambiguous fragment was highlighted. In the *unambiguous* condition, the message to which the unambiguous fragment refers was to be communicated. Note that this fragment was *unambiguous* with respect to the three messages displayed in the experimental setup, though not necessarily in the production data set.[6] As Table 1 shows, on average, the ambiguous fragment had a higher $L_0p(m|u)$ in the *critical* condition than in the *competitor* condition within each scenario (see Table 1). As the ranges show, there was some overlap in the probabilities, but within each scenario, the message in the critical condition was always more likely.

Production cost was implemented with a system of virtual coins. In Experiment 1, participants received a starting balance of 500 coins and the cost for sending a fragment (30 coins) was lower than that for sentences (100 coins). Successful communication was rewarded with 120

[6] In some stimuli, it was also unambiguous in the production data, as the maximal $L_0p(m|u)$ of 1 in Table 1 shows, but this was not always the case.

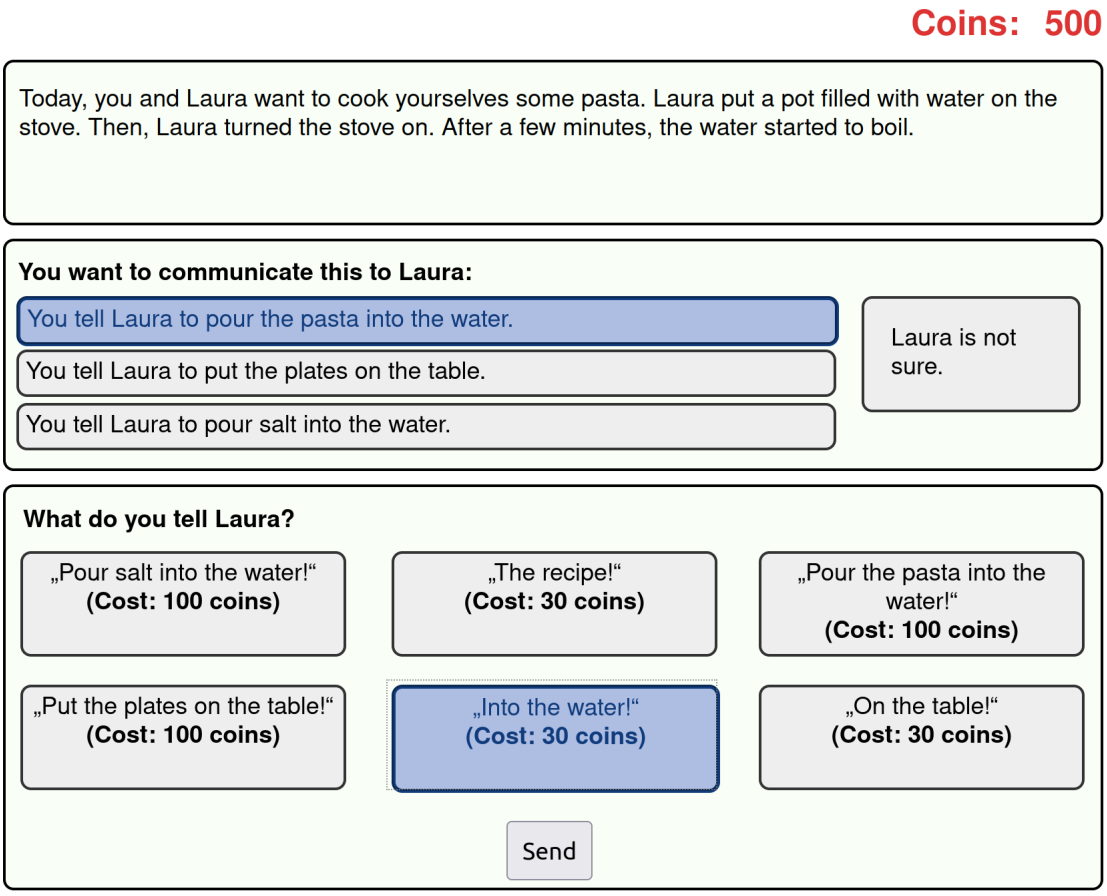

Figure 1: Screenshot of Experiment 1 after revealing the messages and utterances. The experiment was conducted in German, but has been translated here for convenience.

Condition	**Lowest** $p(m\|u)$	**Highest** $p(m\|u)$	**Mean** $p(m\|u)$
critical	0.12	0.69	0.38
competitor	0.03	0.22	0.09
unambiguous	0.15	1.0	0.75

Table 1: Range of $L_0(m|u)$ probabilities and means by conditions

coins.[7] These quantities were set based on the *expected utility* (EU) (Franke 2009) of the utterances in each condition. In the experiment, EU can be calculated as shown in equation 2 by subtracting the cost c from the product of the likelihood of successful communication $L_0p(m|u)$ and the utility $U(m, u)$, i.e. the payoff in case of success.

$$EU(u, m) = L_0p(m|u) \times U(m, u) - c \tag{2}$$

[7]See Section 6 for the cost structure used in Experiment 2.

Since the EU of an (unambiguous) sentence is $EU(u, m) = 1 \times 120 - 100 = 20$ and fragments yield a higher EU if their $L_0p(m|u) > 0.42$, the model predicts an absolute preference for fragments beyond this threshold. Given the data in Table 1, in the range of $L_0p(m|u)$ it should be possible to observe differences in the ratio of fragments produced. On average, the fragment ratio is expected to be highest in the unambiguous condition and lowest in the competitor condition, with the critical condition in between.

4.3. PROCEDURE. Each of the two experiments was completed by 60 self-reported native speakers of German recruited on the crowd-sourcing platform Prolific. The participants were rewarded with £2.67 (the equivalent of the German minimum wage of 12.41€ for the projected duration of 15 minutes). The materials were distributed across three lists, so that each participant saw each of the 15 stimuli in one of three conditions, with each condition appearing equally often. Stimuli were presented in individual fully randomized order. [8]

After giving informed consent by marking a checkbox, reading the instructions and providing their Prolific ID, participants were asked to enter a nickname, which should not be their real name and which would not be published, to make the experiment appear more interactive. They were then asked to wait for the connection to their partner to be established. The waiting time was randomly sampled from an array of durations between 10 and 25 seconds. After this, the participants were notified that a partner had been found, informed about their partner's name (randomly sampled from an array of names), and asked to begin a practice phase of two (Experiment 1) or three (Experiment 2) trials.[9] The setup in the practice phase was identical to the main experiment in terms of the number of messages and utterances, as well as the costs. The only difference was that each of the three fragments referred unambiguously to one of the messages. After the practice phase, the score was reset to the initial amount and participants were informed in advance about this.

Each trial began with a display of the context story only. Participants were told to press the space bar to display each of the other elements: the messages, the highlighted message, and the utterances. This intended to make subjects read all of the messages before seeing which one would be highlighted, which is critical to realize that the ambiguous fragment is ambiguous. After selecting an utterance, participants received feedback on the partner's interpretation. As anticipated above, the partner always maximized $L_0p(m|u)$ when selecting an utterance: For correct sentential utterances and fragments in the critical condition (where the message had a high $L_0p(m|u)$) this results always in correct interpretations. Fragment utterances in the (low $L_0p(m|u)$) competitor condition were never successful, as maximizing $L_0p(m|u)$ led to selecting the more likely interpretation. If communication was successful, a notice "⟨partner⟩ understood you correctly" appeared in the center of the screen, otherwise it read "⟨partner⟩ understood something else". Additionally, the message selected by the fake partner was highlighted in green (success) or red (failure) to indicate what the partner understood. If the participant had selected the attention check fragment, which

[8] The experiment was followed by a questionnaire to evaluate the experiment, which included questions about the participants' affinity to games and self-assessed likelihood of taken risks to investigate potential individual differences. Since there were no theoretically interesting correlations between the answers to these questions and behavioral data, the questionnaire is not discussed further here.

[9] Since there were almost no errors in the practice phase, it was shortened to two trials in Experiment 2.

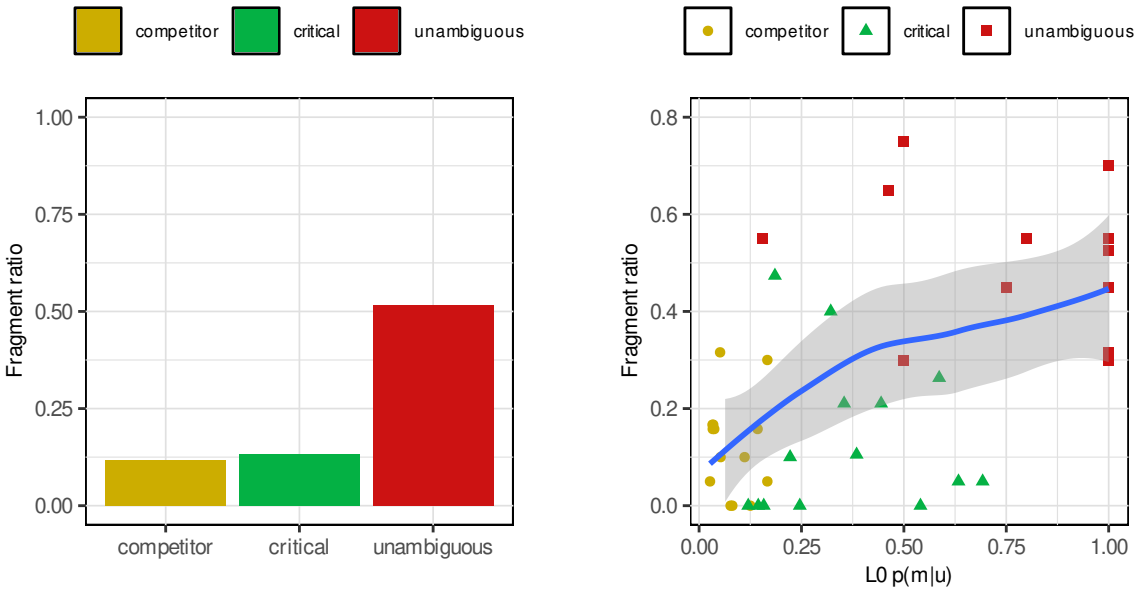

Figure 2: The left facet shows the ratio of fragments across the three conditions, the right facet illustrates the continuous effect of $L_0(m|u)$ on the ratio of fragments with a Loess smooth (span=1).

did not refer to any of the three messages, the text field "⟨partner⟩ is not sure" was highlighted. To simulate the partner's reasoning, a delay was introduced between the participant's sending the message and the feedback. The delay time was randomly sampled from an array of values between 0.8 and 4.4 seconds. Since the participant was told that their partner could see the complete screen throughout the trial, and consequently have read all of the messages and utterances in the meantime, quick responses to the participants' choice seemed natural.

5. Experiment 1. Experiment 1 was conducted using the cost structure described in Section 4.2. Due to an error in preparing the stimulus table, which resulted in the target fragment being unambiguous, this stimulus had to be discarded before the analysis, leaving 14 stimuli for analysis. One participant, who provided more utterances incompatble with the target message (n=8) than compatible ones (n=6), was also excluded from further analysis.

5.1. RESULTS. The participants were very accurate: Only in three trials, the distractor was chosen and in 16 trials an utterance that could not communicate the highlighted message. These 19 trials (2.3% of the data) were excluded before the analyses reported in what follows.

The responses across the three conditions and as a function of $L_0p(m|u)$ are summarized in Figure 2. Overall, the participants had a strong preference for complete sentences: In the ambiguous conditions, the fragment was selected in only about 12-13% of the trials, and even in the unambiguous condition, participants preferred the fragment only in half of the trials. Since the game-theoretic account predicts a gradual effect of an utterance's expected utility on speakers' choices, and there are notable differences in $L_0p(m|u)$ between the scenarios tested (see Table 1), in the statistical analyses, the PROBABILITY of successful communication was used as a continuous predictor in the statistical analyses rather than investigating categorical contrasts between the

three conditions. Descriptively, the right facet of Figure 2 supports the prediction that a higher likelihood of successful communication with fragments increases the fragment ratio.

The data were analyzed with mixed effects logistic regressions (Bates et al. 2015) in R (R Core Team 2024; version 4.4.1) using a backward model selection procedure. The models predicted a binary dependent variable ELLIPSIS (sentence/fragment) from the PROBABILITY ($L_0p(m|u)$) of the fragment and the scaled and centered POSITION of the trial in the experiment. The full model contained the maximal random effects structure supported by the data, which consisted of by-subject and by-item random intercepts and by-item random slopes for PROBABILITY.[10] Starting from this full model, fixed effects that did not significantly improve model fit were subsequently removed. This was determined using Likelihood ratio tests conducted with the `anova` function (R Core Team 2024).

The final model contained only a significant main effect of PROBABILITY ($\chi^2 = 6.5, p < 0.05$), which suggests that, as predicted by the game-theoretic account, the ratio of fragments increases with their $L_0p(m|u)$. This finding is consistent with the game-theoretic approach. However, the left facet of Figure 2 suggests that there may be no difference between the ambiguous (critical and competitor) conditions, despite the higher $L_0p(m|u)$ in the critical condition predicting a higher fragment ratio. Consequently, the PROBABILITY effect might be driven by the unambiguous condition alone. This aligns with the account's predictions, because the fragments in the unambiguous condition had a higher average $L_0p(m|u)$. However, this does not provide genuine evidence for probabilistic rational reasoning, as participants may have simply avoided ambiguity in the experiment. Therefore, I conducted a further regression analysis using only the data for the ambiguous conditions. If participants relied on game-theoretic reasoning, the effect of PROBABILITY should be replicated in this subset. Note that the opposite result would not prove that participants did not rely on game-theoretic reasoning, as they might it and be but less likely to assume risks than predicted by the current model. The analysis followed the same procedure as the main analysis reported above in this section. However, there was no significant main effect of PROBABILITY ($\chi^2 = 0.01, p > 0.9$), failing to confirm the game-theoretic prediction of a gradual effect.

5.2. DISCUSSION. The results of Experiment 1 are in line with the prediction of the game-theoretic account that speakers choose ambiguous fragments more often if the risk of misunderstanding is reduced due to a high $L_0p(m|u)$. This gradual effect was not confirmed when excluding the unambiguous condition, which could have at least two reasons. First, fragments had the highest $L_0p(m|u)$ in this condition, so that the preference for using fragments might simply be stronger. Second, if participants did not rely on game-theoretic reasoning and prior message probabilities, they could simply avoid any expression which is ambiguous in the context of the experiment. While the former explanation would support the game-theoretic account, the second would provide an explanation for the results which does not require probabilistic pragmatic reasoning.

Another result of the experiment is the overall relatively low fragment ratio. Even for some fragments with a $L_0p(m|u)$ of 1 in the unambiguous condition, participants selected the fragments only in 30–50% of the trials. This might be at least partially due to the cost structure, which guaranteed participants a net benefit of 20 coins per trial when using only unambiguous sentences.

[10] Ellipsis ~ Probability * Position + (1 | Subject) + (1 + Probability | Item)

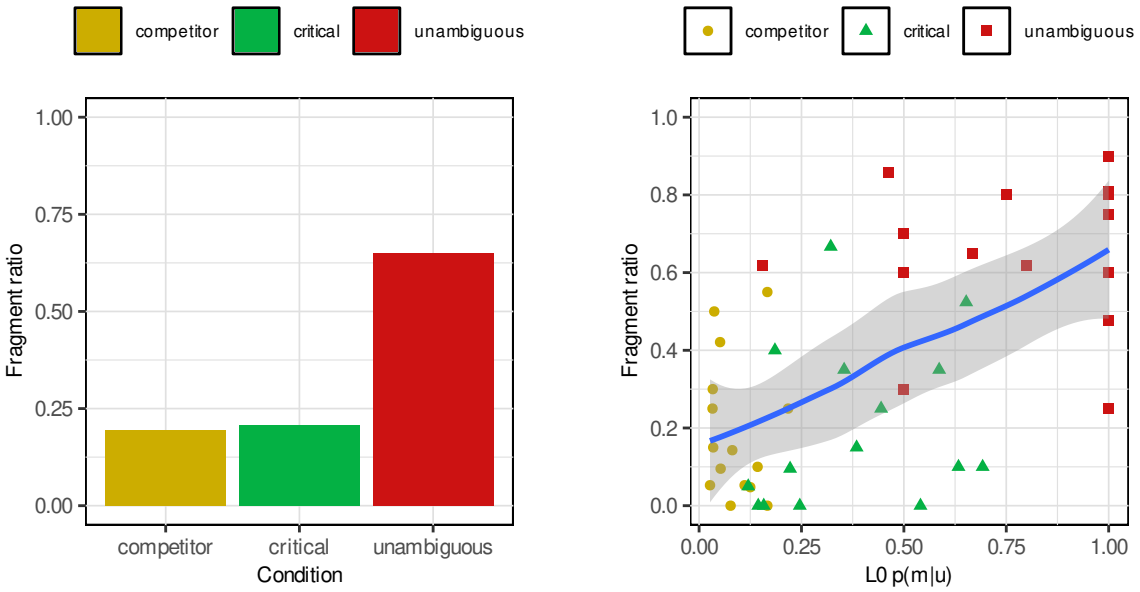

Figure 3: The left facet shows the ratio of fragments across the three conditions, the right facet illustrates the continuous effect of $L_0(m|u)$ on the ratio of fragments with a Loess smooth (span=1).

When looking at the individual participants' data, it turns out that 15 of them correctly completed all trials by selecting only sentences. This suggests that the cost structure used in Experiment 1 biased participants toward adopting a risk-avoiding strategy, which might have obscured gradual differences between the ambiguous conditions.

6. Experiment 2. Experiment 2 addressed the concern that the low fragment ratio in Experiment 1 might have occurred due to a cost structure favoring a risk-avoiding strategy which made about 25% of the participants to select sentences only. To address this, the cost structure was modified to reduce the benefits obtained in Experiment 1 through this strategy.

6.1. MATERIALS AND COST STRUCTURE. The materials used in Experiment 2 were identical to Experiment 1, except that the error resulting in the exclusion of one experimental item was resolved. To make a sentence-only strategy less rewarding, the utterance costs and the starting balance were adjusted: The reward for successful communication was reduced to 100 coins (instead of 120), the cost of sentences increased to 130 (instead of 100) and the cost of fragments increased to 40 (instead of 30). Unlike in Experiment 1, selecting sentences now resulted in a loss of 30 coins per trial instead of a benefit of 20 and the lower starting balance should increase the motivation to keep the score high (even though a negative score had no consequences for the subjects' reward and they were informed about this in advance).

6.2. PROCEDURE. The procedure was identical to Experiment 1. 60 further participants, who did not participate in Experiment 1, were recruited on Prolific and rewarded with £2.67.

6.3. RESULTS. Figures 3 shows the ratio of fragments by condition and as a function of $L_0p(m|u)$. The overall fragment ratio was higher than in Experiment 2 (35.01% instead of 25.32%), which

Predictor	**Est.**	**Std. error**	χ^2	p-value
INTERCEPT	2.14	0.36	17.71	< 0.001
PROBABILITY	-2.56	0.67	9.23	< 0.01
EXPERIMENT	-0.50	0.12	18.02	< 0.001

Table 2: Fixed effects in the final model for the joint analysis of experiments 1 and 2.

was also reflected in participants' individual behavior: Unlike in Experiment 1, only three participants followed a sentence-only strategy, which resulted in a score of -150. This suggests that the smaller but secure reward that this strategy returned in Experiment 1 was (at least partially) responsible for the higher relatively low fragment usage in Experiment 1). Besides the higher fragment ratio, the pattern looks similar to Experiment 1: Fragment ratio seems to increase as a function of $L_0p(m|u)$, but there is no obvious difference between the ambiguous conditions.

I first conducted an analysis in parallel with Experiment 1 using mixed effects logistic regressions (Bates et al. 2015; version 1.1-35.3) in R (R Core Team 2024; version 4.4.1). The full model contained by-subject and by-item random intercepts and by-item slopes for PROBABILITY, as well as fixed effects for PROBABILITY, POSITION (scaled and centered) and their interaction. The overall pattern was identical to Experiment 1: In the analysis of the complete data set, there was a significant main effect of PROBABILITY ($\chi^2 = 11.04, p < 0.001$). As in Experiment 1, this effect was not significant within the ambiguous conditions only ($\chi^2 = 0.4, p > 0.5$).

To quantify the difference between the experiments, I then conducted a joint analysis of the data from both studies. The procedure was identical to the regression analyses reported above, except that I included an additional predictor EXPERIMENT (sum-coded, $exp.1 = -0.5$, $exp.2 = 0.5$) and its interactions with PROBABILITY (numeric) and POSITION (numeric, scaled and centered). A significant main effect of EXPERIMENT would statistically confirm the higher fragment ratio in Experiment 2 and an EXPERIMENT:PROBABILITY interaction would show whether the effect of PROBABILITY is also stronger. The full model contained main effects for all three predictors and all interactions between them, as well as by-items random intercepts and random slopes for PROBABILITY. The final model (See Table 2) contains a main effect of PROBABILITY, which had been also found in the analysis of the individual experiments ($\chi^2 = 9.23, p < 0.01$), a main effect of EXPERIMENT ($\chi^2 = 18.02, p < 0.001$). The latter effect confirms the intuition that the fragment ratio is higher in Experiment 2. The interaction between both predictors is not significant ($\chi^2 = 2.14, p > 0.1$).

6.4. DISCUSSION. Experiment 2 investigated whether the low fragment ratio in Experiment 1 was (partially) due to an overall bias toward sentences in order to avoid risks. This could have obscured gradual differences between the ambiguous conditions, which would have provided genuine evidence for pragmatic reasoning in fragment usage. The significant difference in fragment ratio between both experiments, confirmed by the joint analysis, suggests that the cost structure had an effect: If fragments yield a higher potential benefit, their ratio is increased, even if the likelihood of communicative success remains constant. However, it is also clear that the stronger preference for sentences in Experiment 1 was not the only cause for the lack of a gradual effect within the ambiguous conditions, since this effect was also not found in Experiment 2.

7. General discussion. I conducted two pseudo-interactive utterance selection experiments to investigate whether the usage of fragments follows the predictions of a game-theoretic account. Unlike previous studies in the field, I did not use tightly restricted and balanced sets of utterances and messages; instead, I constructed the stimuli based on a diverse, crowd-sourced data set, with a much larger variety of messages and utterances, as well as unbalanced priors over messages.

The results are overall in line with the trade-off between reducing production cost and maximizing the likelihood of successful communication predicted by the game-theoretic account: In the main analyses of both experiments, participants preferred fragments more often when the model predicted a high likelihood of getting the message across. They also selected more fragments in Experiment 2, where the modified cost structure increased the gain in efficiency as compared to Experiment 1. However, the effect observed in the complete data set was not found when examining only the data from the ambiguous conditions. Experiment 2 investigated whether it had been obscured by the bias toward using sentences in Experiment 1, but this was not the case. Therefore, at this point it cannot be ruled out that participants simply avoided ambiguity in the experimental setup, which does not require pragmatic reasoning. A possible reason for the absence of the gradual effect might be the method used to collect the production data set. Since participants provided only a single most likely utterance, less likely utterances are probably underrepresented in the data set. This could result in low $L_0p(m|u)$ messages being relatively predictable, which may have biased subjects to prefer sentences more often.

Taken together, the experiments show how a game-theoretic account of language production and interpretation can also be applied to more realistic and diverse communication situations than the previous tightly controlled studies on reference. They also suggest that the choice between an elliptical and a sentential utterance depends on a trade-off between production cost and the likelihood of communicative success, which can be explicitly captured by a game-theoretic model of pragmatic inference. Future research could explore the extent to which this reasoning is required only for interpreting discourse-initial fragments tested in the experiments presented here, or whether it also extends to other types of ellipsis.

References

Arregui, Ana, Charles Clifton, Lyn Frazier & Keir Moulton. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* 55(2). 232–246. 10.1016/j.jml.2006.02.005.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. 10.18637/jss.v067.i01.

Bergen, Leon & Noah Goodman. 2015. The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science* 7(2). 336–350. 10.1111/tops.12144.

Burnett, Heather. 2017. Sociolinguistic interaction and identity construction: The view from game-theoretic pragmatics. *Journal of Sociolinguistics* 21(2). 238–271. 10.1111/josl.12229.

Frank, Michael. C. & Noah. D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336(6084). 998–998. 10.1126/science.1218633.

Franke, Michael. 2009. *Signal to act: Game theory in pragmatics*: Universiteit van Amsterdam dissertation.

Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative investigations: The form, meaning, and use of English interrogatives*. Stanford, California: CSLI Publications.

Grice, H. Paul. 1975. Logic and conversation. In P. Cole & J.L. Morgan (eds.), *Syntax and semantics*, vol. 3, 41–58. New York: Academic Press.

Lemke, Robin. 2021. *Experimental investigations on the syntax and usage of fragments* (Open Germanic Linguistics 1). Berlin: Language Science Press.

Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27(6). 661–738. 10.1007/s10988-005-7378-3.

Morgan, Jerry. 1973. Sentence fragments and the notion 'sentence'. In Braj B. Kachru, Robert Lees, Yakov Malkiel, Angelina Pietrangeli & Sol Saporta (eds.), *Issues in linguistics. Papers in honor of Henry and Renée Kahane*, 719–751. Urbana: University of Illionois Press.

Ott, Dennis & Volker Struckmeier. 2016. Deletion in clausal ellipsis: Remnants in the middle field. *UPenn Working Papers in Linguistics* 22(1). 225–234.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Reich, Ingo. 2007. Toward a uniform analysis of short answers and gapping. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure, meaning and form*, 467–484. Amsterdam: John Benjamins. 10.1075/la.100.25rei.

Rohde, Hannah, Scott Seyfarth, Brady Clark, Gerhard Jaeger & Stefan Kaufmann. 2012. Communicating with cost-based implicature: A game-theoretic approach to ambiguity. In *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue*, 107–116.

Schäfer, Lisa, Robin Lemke, Heiner Drenhaus & Ingo Reich. 2021. The role of UID for the usage of Verb Phrase Ellipsis: Psycholinguistic evidence from length and context effects. *Frontiers in Psychology* 12. 10.3389/fpsyg.2021.661087.

Sikos, Les, Noortje J. Venhuizen, Heiner Drenhaus & Matthew W. Crocker. 2021. Reevaluating pragmatic reasoning in language games. *PLOS ONE* 16(3). e0248388. 10.1371/journal.pone.0248388.

Wanzare, Lilian D. A., Alessandra Zarcone, Stefan Thater & Manfred Pinkal. 2016. DeScript: A crowdsourced corpus for the acquisition of high-quality script knowledge. In *Proceedings of LREC 2016*, 3494–3501. Portoroz, Slovenia.

Weir, Andrew. 2014. *Fragments and clausal ellipsis*: University of Massachusetts Amherst dissertation.