Context affects the comprehension of implicit arguments: Evidence from the maze task

Tobias Ungerer (tobias.ungerer@concordia.ca)
Department of Psychology, 7141 Sherbrooke Street West
Montreal, QC H4B 1R6 Canada

Roberto G. de Almeida (roberto.dealmeida@concordia.ca)
Department of Psychology, 7141 Sherbrooke Street West
Montreal, QC H4B 1R6 Canada

Abstract

Linguistic arguments can be either explicitly realized (*After she phoned him, …*) or left implicit (*After she phoned [Ø], …*). In production, the choice between these options is thought to depend on the contextual predictability of the implied referent. We investigated whether different contextual referents (single vs. multiple vs. underspecified) also affect the *comprehension* of implicit arguments, using the "maze" variant of self-paced reading. Our results suggest that, rather than predictability, other context-dependent pragmatic effects, such as the perceived genericness of actions, may influence how speakers comprehend implicitly encoded information.

Keywords: implicit arguments; context effects; predictability; informativity; comprehension; maze task

Introduction

Communication is subject to competing pressures, such as the desire for clarity and economy (Leech, 1983). While clarity requires that speakers specify all participants involved in an event (e.g., the subject, the object, etc.), economy may induce speakers to omit such information if it is sufficiently predictable from the context. In (1), for example, the verb *phone* occurs without an explicit object argument, but the context implies that the recipient of the phone call is Fred. Such "implicit arguments" (Larson, 1988) are also known as "unarticulated constituents" (Perry, 1986) or "null complements" (Fillmore, 1986).

(1) *Sandra decided to call Fred. After she phoned [Ø], Sandra went outside for a walk.*

While contextual predictability is assumed to be a key factor determining the nature of implicit arguments, little experimental research has addressed how its effects bear out during real-time language processing. Some recent work suggests that predictability (or informativity) affects the omission of arguments in *production* (Kline, Gibson, & Schulz, 2018; Kline, Schulz, & Gibson, 2017). Meanwhile, it is unclear what role contextual predictability plays during the *comprehension* of implicit arguments, which may happen much more rapidly and rely on automatic processing rather than conscious message planning. In other words: Do contextual factors determine how speakers infer implicit information during incremental on-line comprehension?

In two experiments, we used the "maze" variant of self-paced reading (Forster, Guerrera, & Elliot, 2009) to investigate how the presence of different contextual referents affects the comprehension of implicit arguments. Our goal was, first, to assess whether contextual predictability affects processing in this modality, and second, to identify other context factors that may influence speakers' expectations about the encoding of upcoming linguistic arguments.

Contextual influences on the processing of implicit arguments

Linguistic arguments may be left implicit for different reasons. Here, we focus on so-called *definite* null complements (Fillmore, 1986), where the missing argument denotes a specific entity that is recoverable from the discourse context, such as *Fred* in (1). We leave aside cases of *indefinite* null complements in which no specific referent is available, as in (2), where *eat* is used in a habitual sense and its referent is therefore irrelevant within the discourse.

(2) *I eat [Ø] three times a day.*

Numerous theoretical studies have highlighted the role of the discourse-pragmatic context in licensing (definite) implicit arguments (Fillmore, 1986; Goldberg, 2001; Lambrecht & Lemoine, 2005). In particular, it is assumed that implicit arguments are more felicitous the more easily the implied referent can be predicted from the context. Evidence for this role of contextual predictability comes from Kline, Schulz, and Gibson's (2017) production study. They first showed participants context pictures that depicted different numbers of potential agents (e.g., people) and patients (e.g., animals). Participants were then asked to describe events involving an agent and a patient, but they could only use two words (including the verb). The authors found that participants were more likely to mention the agent and omit the patient (e.g., *John feeds [Ø]*) the more agents and the fewer patients had been included in the context pictures. In other words: The more predictable the identity of the patient was (since there were fewer options), the more likely participants were to omit it from their descriptions (see Kline, Gibson, & Schulz, 2018, for similar results with young children).

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

However, it is less clear how contextual predictability affects the *comprehension* of implicit arguments. Some preliminary evidence comes from Besserman and Kaiser's (2016) self-paced reading study, in which participants read sentences like (3).

(3) *While the man hunted [Ø] (,) the deer ran into the woods near the house.*

In this example, *hunted* is used with an implicit object, as highlighted by the presence of a comma after the verb. If this comma is missing, however, participants are likely to initially interpret *the deer* as the object of *hunted*, an analysis that they subsequently need to revise when they encounter the word *ran* – a classic garden-path effect (Frazier & Rayner, 1982).

In Besserman and Kaiser's study, the garden-path-inducing stimuli were preceded by a context sentence that either introduced the implied referent (*the deer*) or did not. The authors' rationale was that, if contextual predictability facilitates the comprehension of implicit arguments, then pre-activating the implied referent should alleviate the garden path in the target sentence. Contrary to this prediction, however, the authors found no statistically reliable effect of context on reading times at the disambiguating target word (even though there was a numerical tendency in the expected direction at one of the spillover words).

Davis and van Schijndel (2020) replicated Besserman and Kaiser's (2016) design, using a natural language processing (NLP) model rather than human participants. Their results were similar, in that the contextual predictability of the implied referent did not affect the model's surprisal (as a proxy for processing cost) at the target disambiguation point.

Taken together, these findings suggests that context does not influence the comprehension of implicit arguments in the same way as in production. Nevertheless, the absence of context effects in these studies could also be due to their methodological limitations. In particular, traditional self-paced reading suffers from the fact that the effects are often not localized to the target word but spill over into neighboring regions (Witzel, Witzel, & Forster, 2012). As for NLP studies, while their results can predict many aspects of human performance, recent work still points to relevant differences between the two, especially where context effects are concerned (Cai, Haslett, Duan, Wang, & Pickering, 2023).

In our study, we addressed these concerns by using an alternative comprehension method that may overcome the limitations of traditional self-paced reading. In addition, we investigated more variable context effects than has been previously done. Specifically, the above studies only tested how one specific context feature – the presence or absence of the implied referent – affects the comprehension of implicit arguments. In contrast, we used further manipulations to explore how speakers' processing is influenced by the presence of multiple and underspecified referents.

The present study

In two experiments, we investigated how aspects of the referential context affect the comprehension of implicit arguments. Participants read target sentences like (4), containing an implicit object argument, such as *drank [Ø]*, which was either followed by a comma or not.

(4) *While Andrei drank [Ø] (,) the music played loudly from the speakers*.

In the version without a comma, speakers were likely to interpret the determiner after the verb (*the*) as introducing a direct object (i.e., *drank the …*). The subsequent noun (*music*), however, was semantically incompatible with this interpretation, thus alerting participants to the presence of an implicit argument and forcing them to reanalyze the sentence.[1] We expected that this would result in higher processing difficulty, thus giving rise to a garden-path effect. Our goal was to examine how much the severity of the garden path is modulated by different context manipulations.

In Experiment 1, we investigated how the contextual predictability of available referents affects speakers' comprehension of implicit arguments. The targets were preceded by context passages that introduced either a single referent or two referents. We predicted the following: When only a single referent (e.g., *beer*) is contextually available, then the implied referent in the following target (e.g., the referent of *drank* in (4)) is highly predictable. As a result, speakers should find it easier to reanalyze the target in the version without a comma, thus giving rise to a weaker garden path effect. On the other hand, if the context introduces two referents (*beer* and *whiskey*), then it is more difficult to predict the identity of the implied referent in the target, so the garden path effect should be stronger. As a result, we expected higher processing difficulty (as manifested, for instance, by longer response times) at the disambiguation point of the targets (e.g., *music* in (4)) in two-referent contexts compared with single-referent contexts.

In Experiment 2, we introduced further context manipulations which left the referents underspecified (e.g., describing them as *something (particular)*). We discuss the predictions for Experiment 2 in the respective section below.

To assess participants' comprehension, we used the "maze task" (Forster, Guerrera, & Elliot, 2009). In this task, participants read sentences word-by-word while having to choose between two options at every step: a correct sentence continuation and an incorrect distractor. The maze task has several advantages over traditional self-paced reading: It appears to index well-known processing effects more robustly (Boyce, Futrell, & Levy, 2020; Witzel, Witzel, & Forster, 2012); it largely avoids spillover effects (Boyce & Levy, 2023); and its higher task demands help ensure that participants stay attentive while reading.

[1] Our materials thus differed from Besserman and Kaiser's (2016), where reanalysis was only triggered at the verb of the main clause (e.g., at *ran* in (3)). We piloted a version of Experiment 1 with stimuli more similar to theirs, but the results were less clear, which is why we adopted the present design.

Experiment 1

Method

Participants We recruited 40 participants via the crowdsourcing platform Prolific (www.prolific.com). Participation was restricted to individuals who reported English as their first and primary language, who currently lived in the UK, US, or Canada and had resided there for at least two years, who did not declare any language-related disorders or dyslexia, and who had an approval rating of 95% or higher. Participants received £3 as compensation.

Materials We created 24 text passages, each of which comprised two context sentences and a target sentence. There were four conditions, as illustrated in Table 1.

Table 1: Examples of the four conditions in Experiment 1 (differences in bold; disambiguating word underlined).

Condition	Context	Target
Single referent; garden path	Andrei went to a pub where he was a regular. The bartender immediately served him **a beer**.	While Andrei **drank** the <u>music</u> played loudly from the speakers.
Two referents; garden path	Andrei went to a pub where he was a regular. The bartender immediately served him **two things: a beer and a whiskey**.	While Andrei **drank** the <u>music</u> played loudly from the speakers.
Single referent; no garden path	Andrei went to a pub where he was a regular. The bartender immediately served him **a beer**.	While Andrei **drank,** the <u>music</u> played loudly from the speakers.
Single referent; explicit object	Andrei went to a pub where he was a regular. The bartender immediately served him **a beer**.	While Andrei **drank the beer** the <u>music</u> played loudly from the speakers.

The first two conditions introduced either a single referent (*a beer*) or two referents (*a beer and a whiskey*) in the context. The targets contained an implicit argument (*drank [Ø]*) followed by no comma, thus likely inducing a garden path.[2] The third condition was identical to the first, except that the targets contained a comma, intended to disambiguate the sentences and avoid the garden path. The fourth condition was an additional control in which the targets contained an explicit object (*the beer*) instead of an implicit one. We included this condition so that participants could not learn to predict that the targets always contained implicit arguments.[3]

For the purposes of the maze task, we created distractors for each word of our target sentences, using Boyce, Futrell, and Levy's (2020) "Auto-maze." Based on a large language model, this tool automatically generates distractors that resemble the target words in length and frequency, but which are contextually inappropriate (i.e., causing high surprisal).

We divided the stimuli over four lists such that participants only saw one version of each passage, in random order. We added 12 fillers that contained unrelated linguistic structures.

Procedure The experiment was conducted on the PCIbex platform (Zehr & Schwarz, 2018). Participants completed two practice trials before starting the experiment. In each trial, participants first read the two context sentences, presented together as a paragraph, and then pressed the space bar. On the next screen, they saw the target sentence presented word-by-word in the maze format, with each word displayed next to a distractor in the center of the screen (their positions were randomized). Participants selected the correct sentence continuation by pressing "e" for the word on the left and "i" for the word on the right. When participants chose the incorrect option, an error message was displayed for 1 second, after which they were allowed to correct their response. 50% of trials were followed by a comprehension question about the content of the context sentences.

Data analysis We excluded two participants who failed to reach 80% accuracy in the maze task and/or 70% accuracy in the comprehension questions. They were replaced by two new participants. We also excluded data from two items where the text passages contained errors.

Using R (R Core Team, 2023), we then analyzed participants' response times and accuracy in the maze task at the disambiguating word of the target sentences (e.g., *music* in Table 1).[4] For the response times, we only considered words where participants had chosen the correct maze option (thus excluding 3.7% of the data). We also removed response times below 100 ms and above 5000 ms (0.8%) as well as response times that were more than 2.5 standard deviations away from each participant's mean (3.1%). As preliminary modeling indicated that the assumption of normality of

[2] We selected verbs that are biased towards transitive uses (i.e., with a direct object), thus making the garden path more likely. Most verbs were taken from previous garden-path studies (Grodner, Gibson, Argaman, & Babyonyshev, 2003; Sturt, Pickering, & Crocker, 1999), which include norming data about the verb bias. Moreover, we made sure that the transitive use is listed first within each verb's entry in the Merriam-Webster dictionary.

[3] A reviewer asks why we did not run a full factorial design, crossing number of referents with presence/absence of comma. Note, however, that our interest was primarily in the contrast between the two garden-path conditions, and that the comparison with the comma-containing condition served as a (logically independent) replication of the garden-path effect.

[4] We also analyzed the following two words as a potential spillover region but found no significant differences.

residuals was violated, we log-transformed the response times to render their distribution more normal (Baayen & Milin, 2010).

We then fitted a linear mixed model for response times and a logistic mixed model for accuracy, using the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). The fixed effect was "structure type," instantiated by the four conditions of our text passages. We added random intercepts for subject and item, while random by-participant slopes for structure type did not converge. We compared each model to a null model without structure type via a likelihood ratio test. We then conducted pairwise post-hoc comparisons between the four conditions with the package *emmeans* (Lenth, 2023), using Tukey adjustments for multiple comparisons.

Results

For response times, the comparison with the null model indicated an overall effect of structure type ($\chi^2(3) = 31.02$, $p < .001$). Back-transformed model estimates for the four conditions are depicted in Figure 1.

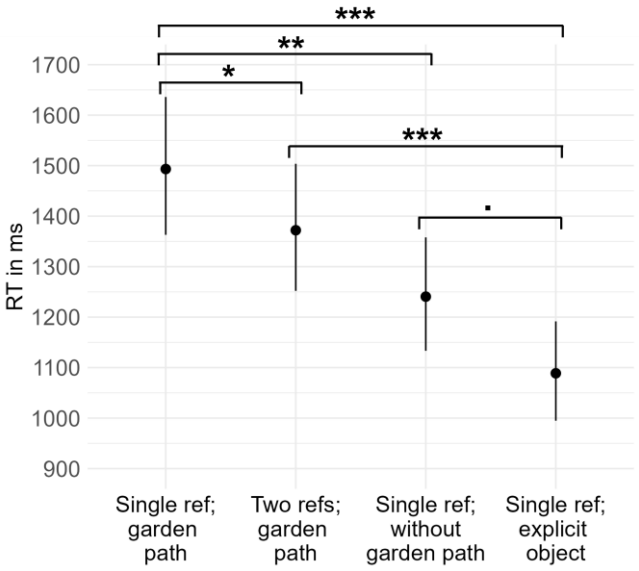

Figure 1: Response times at the disambiguating word in Experiment 1 (error bars mark 95% confidence intervals).

Post-hoc comparisons indicated that there was a statistically significant difference between the two garden path conditions, with longer response times in single-referent contexts than in two-referent contexts ($t = 2.82$, $p = .03$). The single-referent *with* garden path condition also produced longer response times than the single-referent *without* garden path condition ($t = 3.36$, $p = .007$). Finally, the control condition that contained an explicit object in the targets yielded shorter response times than the two garden path conditions (vs. single referent: $t = 5.73$, $p < .001$; vs. two referents: $t = 5.73$, $p < .001$), and marginally shorter responses than the single-referent without garden path condition ($t = 2.384$, $p = .09$).

Accuracy in the maze task was high (> 90%) across all conditions. While the model comparison indicated an overall effect of structure type on accuracy ($\chi^2(3) = 8.68$, $p = .03$), none of the pairwise comparisons between the four conditions reached significance (all *p*s > .05).

Discussion

A first finding of Experiment 1 was the expected garden path effect (Frazier & Rayner, 1982): Target sentences without a comma yielded longer response times at the disambiguating word than targets with a comma (both in single-referent contexts), suggesting that participants initially misinterpreted the no-comma sentences as containing an explicit object. We take this as validation of the maze task as a means of studying garden-path effects in comprehension.

Our critical context manipulation also had an effect on speakers' processing of garden path sentences, but not in the expected direction: When the context introduced a single referent, speakers took longer to recover from the garden path in the targets than when the context contained two referents. This suggests that when only one referent was available (e.g., *beer*), speakers were *less* likely to construe the target verb (e.g., *drank*) as involving an implicit object than when two referents were available (e.g., *beer* and *whiskey*). This result runs counter to our expectation that the implied referent in the target should be more predictable in the case of a single contextual referent, and thus easier to process.

Rather than predictability, other context-related effects may therefore have driven our results. One possibility is that our context manipulation affected the perceived *genericness* of the target actions. Specifically, participants may have been more likely to interpret an action that involves multiple possible referents as generic, and thus as not requiring an explicit object. For example, when someone drinks several things, it might be more natural to construe this situation as a generic act of 'drinking' than when they consume a single specific drink. In two-referent contexts, participants may have thus found it easier to recover from their initial misanalysis of the garden-path target.

In Experiment 2, we further investigated this possible effect of genericness, while also manipulating the role of contextual predictability.

Experiment 2

In Experiment 2, we used two additional context manipulations to examine the effects of genericness and contextual predictability on the comprehension of implicit arguments. In one condition, the context introduced an underspecified referent (described merely as *something* or *someone*). This, we reasoned, should induce participants to interpret the action in the target sentence (e.g., *drank* in (4)) as generic (in the sense of 'he drank, but it does not matter what'). As a result, speakers should find it easier to process the implicit argument, thus leading to a lower processing cost than when the context contains a single specific referent (e.g., *beer*).

In the other condition, the context contained an underspecified but clearly individuated referent (e.g., *something particular*). This should make it more difficult for participants to interpret the action in the target generically, thus leading to similar processing difficulty as when the context contains a specific referent (*beer*).

On the other hand, if contextual predictability were to have an effect (contrary to the results of Experiment 1), then both types of contexts with underspecified referents should make it more difficult to predict the implied referent in the target. In this case, we would expect higher processing difficulty in both these conditions, compared with contexts that specify the exact nature of the referent.

Method

Participants We recruited 40 participants via Prolific who had not participated in Experiment 1. The restrictions for participation and the compensation remained the same.

Materials We used the same text passages as in Experiment 1, while modifying two of the context conditions, as illustrated in Table 2.

Table 2: Examples of the four conditions in Experiment 2 (differences in bold; disambiguating word underlined).

Condition	Context 1	Context 2	Target
Single referent garden path	Andrei went to a pub where he was a regular.	The bartender immediately served him **a beer**.	While Andrei **drank** the music played loudly from the speakers.
Under-specified referent garden path	Andrei went to a pub where he was a regular.	The bartender immediately served him **something**.	While Andrei **drank** the music played loudly from the speakers.
Indivi-duated referent garden path	Andrei went to a pub where he was a regular.	The bartender immediately served him **something particular**.	While Andrei **drank** the music played loudly from the speakers.
Single referent explicit object	Andrei went to a pub where he was a regular.	The bartender immediately served him **a beer**.	While Andrei **drank the beer** the music played loudly from the speakers.

The "Single referent; garden path" condition and the "Single referent; explicit object" (= control) condition remained the same. We included two new conditions, in which the context either contained an underspecified referent (*something*/ *someone*) or an underspecified but individuated referent (e.g., *something particular*, *someone important*). Both conditions were combined with garden-path-inducing targets (i.e., without commas). We did not include targets with a comma because Experiment 1 had already established the garden path effect, and our interest was in comparing relative differences between the other conditions.

Procedure The procedure was identical to Experiment 1.

Data analysis We excluded one participant who failed to reach 70% accuracy in the comprehension questions and replaced them with a new participant.

For response times, we excluded words for which participants had selected the incorrect maze option (3.2% of the data) as well as response times below 100 ms and above 5000 ms (1.1% of the data) and datapoints that were more than 2.5 standard deviations away from each participant's mean (2.8% of the data). We log-transformed the remaining data to render their distribution more normal.

The procedures for modeling response times and accuracy at the disambiguating word were identical to Experiment 1.

Results

For response times, the comparison with the null model indicated an overall effect of structure type ($\chi^2(3) = 27.08$, $p < .001$). Back-transformed model estimates are depicted in Figure 2.

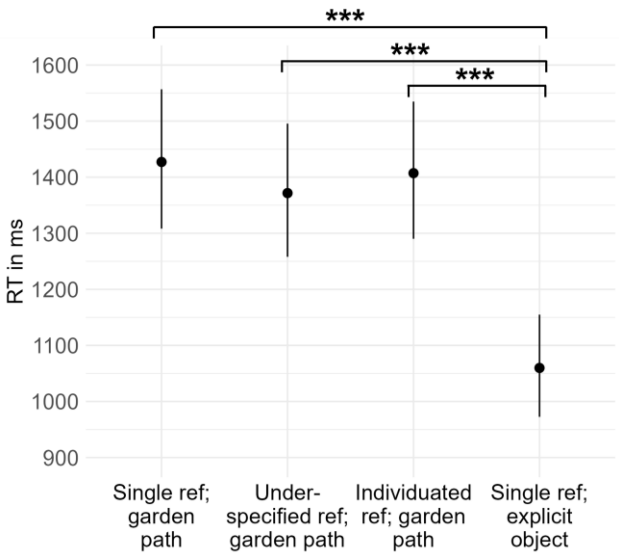

Figure 2: Response times at the disambiguating word in Experiment 2 (error bars mark 95% confidence intervals).

Post-hoc comparisons suggested that response times in the control condition that contained explicit objects in the targets were shorter than in the other three conditions (all ps < .001). However, none of the other conditions differed significantly (all ps > .01).

Accuracy was high (> 90%) across all conditions. While the model comparison indicated an effect of structure type

that approached significance ($\chi^2(3) = 6.95$, $p = .07$), none of the pairwise comparisons between conditions were significant (all ps > .01).

Discussion

Experiment 2 did not provide evidence that the presence of different contextual referents (single specific, underspecified, and underspecified but individuated) had differential effects on participants' comprehension of implicit arguments. The fact that response times were comparable across all three garden-path conditions does not support our prediction that underspecified (but non-individuated) referents would trigger a more generic interpretation of the target verbs, thus facilitating speakers' comprehension of the implicit argument. Moreover, and similar to Experiment 1, we found no effect of contextual predictability in that underspecified and thus less predictable referents did not exacerbate the garden path effect.

A possible explanation for these results is that specific referents (e.g., *a beer*) and underspecified referents (e.g., *something (particular)*) function in similar ways as discourse-old (i.e., informationally given) contextual "fillers" which can be subsequently reactivated to fill the (implicit) argument slot in, for instance, *drank* ___. That is, even underspecified expressions might establish the identity of the contextual referent to a sufficient enough degree for this referent to resemble the behavior of a fully specified entity. This contrasts with Experiment 1, where the presence of multiple referents provided speakers with additional contextual information that presumably facilitated their interpretation of the target action as generic.

General discussion

In two experiments, we investigated how speakers' comprehension of implicit arguments is affected by contextual information about the implied referent. Our results do not provide evidence that speakers' processing is influenced by the degree to which the referent is contextually predictable. While this aligns with the results of previous comprehension studies (Besserman & Kaiser, 2016; Davis & van Schijndel, 2020), it contrasts with evidence from production according to which predictable arguments are more often left implicit (Kline, Gibson, & Schulz, 2018; Kline, Schulz, & Gibson, 2017). Our results thus suggest that contextual information plays differential roles during the comprehension and production of implicit arguments. The existence of such differences is plausible: For example, while speakers may actively try to avoid mentioning highly predictable (and thus less informative) arguments in production, this does not mean that they find it difficult to *comprehend* predictable information (see Rohde, Futrell, & Lucas, 2021, for discussion).[5]

Our results also contrast with other comprehension scenarios in which contextual predictability has been shown to affect speakers' processing. For example, restrictive relative clauses and other noun phrase modifiers are comprehended more easily when the context contains multiple competing referents rather than a single possible referent (Altmann, Garnham, & Dennis, 1992; Grodner, Gibson, & Watson, 2005; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). But while the discourse function of restrictive relative clauses and modifiers is specifically to *distinguish* between competing referents, the function of explicit and implicit arguments is merely to refer to (new or already known) entities. These pragmatic differences may explain why contextual predictability affects the former phenomenon more strongly than the latter.

Instead of predictability, our results suggest that *other* contextual factors may shape the comprehension of implicit arguments. Specifically, in Experiment 1, implicit arguments (e.g., *drank [Ø]*) were processed more quickly when the context had introduced multiple possible referents (e.g., *beer* and *whiskey*) rather than a single referent (e.g., *beer*). This indicates that speakers comprehended implicit arguments more easily even when their referents were not fully predictable. Meanwhile, the results of Experiment 2 indicated that underspecified referents, whether non-individuated (e.g., *something*) or individuated (*e.g., something particular*), did not have a similar facilitatory effect on speakers' processing.

These findings suggest that the nature of the referents that form part of the discourse context or conversational "scoreboard" (Lewis, 1979) affects the comprehension of implicit arguments. When the set of referents consists of multiple members (e.g., {*beer, whiskey*}), speakers seem to have a stronger expectation that subsequent utterances invoke these arguments implicitly rather than explicitly. A feasible explanation for this finding is that actions which involve multiple referents are more likely to be construed as generic, thus obviating the need to make their object arguments explicit. On the other hand, contexts that only specify a single-member set of referents (e.g., {*beer*} or {*something*}) do not provide the additional information that would invite such a generic interpretation.

Our study suggests that the comprehension of implicit arguments – and, perhaps, other context-sensitive structures – can be affected by diverse context factors, which are distinct from the predictability effects that have been the focus of much previous research (e.g., Kuperberg & Jaeger, 2016; Rohde, Futrell, & Lucas, 2021). These factors could be further investigated by adapting our experimental design, which combines the maze task with naturalistic context passages in an easily implementable online setting. We are confident that a wider range of context manipulations, when applied to more diverse phenomena, can shed light on the subtle ways in which context guides speakers' processing.

[5] As a reviewer points out, another difference between Kline and colleagues' studies and ours is that they introduced several possible referents in the context whose relationship to the event was left unspecified, while our referents were introduced as necessary participants of the event. Follow-up work will need to address to what extent these referent roles can explain the varying results.

References

Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language*, *31*(5), 685–712.

Baayen, H. R., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1).

Besserman, A., & Kaiser, E. (2016). The effects of discourse cues on garden path processing. *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 877–882). Austin, TX: Cognitive Science Society.

Boyce, V., Futrell, R., & Levy, R. P. (2020). Maze made easy: Better and easier measurement of incremental processing difficulty. *Journal of Memory and Language*, *111*, 104082.

Boyce, V., & Levy, R. (2023). A-maze of Natural Stories: Comprehension and surprisal in the Maze task. *Glossa Psycholinguistics*, *2*(1), 1–34.

Cai, Z. G., Haslett, D. A., Duan, X., Wang, S., & Pickering, M. J. (2023). *Does ChatGPT resemble humans in language use?* arXiv. https://doi.org/10.48550/arXiv.2303.08014

Davis, F., & van Schijndel, M. (2020). Interaction with context during recurrent neural network sentence processing. *Proceedings of the 42th Annual Conference of the Cognitive Science Society* (pp. 2744–2750). Austin, TX: Cognitive Science Society.

Fillmore, C. J. (1986). Pragmatically controlled zero anaphora. *Annual Meeting of the Berkeley Linguistics Society*, *12*, 95–107.

Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*, *41*(1), 163–171.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210.

Goldberg, A. E. (2001). Patient arguments of causative verbs can be omitted: The role of information structure in argument distribution. *Language Sciences*, *23*(4), 503–524.

Grodner, D., Gibson, E., Argaman, V., & Babyonyshev, M. (2003). Against repair-based reanalysis in sentence comprehension. *Journal of Psycholinguistic Research*, *32*(2), 141–166.

Grodner, D., Gibson, E., & Watson, D. (2005). The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. *Cognition*, *95*(3), 275–296.

Kline, M., Gibson, E., & Schulz, L. (2018). Young children choose informative referring expressions to describe the agents and patients of transitive events. *PsyArXiv*. https://doi.org/10.31234/osf.io/r6mwb

Kline, M., Schulz, L., & Gibson, E. (2017). Partial truths: Adults choose to mention agents and patients in proportion to informativity, Even if it doesn't fully disambiguate the message. *Open Mind*, *2*(1), 1–13.

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

Lambrecht, K., & Lemoine, K. (2005). Definite null objects in (spoken) French: A Construction-Grammar account. In M. Fried & H. C. Boas (Eds.), *Grammatical Constructions: Back to the roots*. Amsterdam & Philadelphia: John Benjamins.

Larson, R. K. (1988). Implicit arguments in situation semantics. *Linguistics and Philosophy*, *11*(2), 169–201.

Leech, G. (1983). *Principles of pragmatics*. London: Longman.

Lenth, R. (2023). *Emmeans: Estimated marginal means, aka least-squares means. R package version 1.8.5*. https://CRAN.R-project.org/package=emmeans

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*(1), 339–359.

Perry, J. (1986). Thought without representation. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *60*, 263–283.

R Core Team. (2023). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Rohde, H., Futrell, R., & Lucas, C. G. (2021). What's new? A comprehension bias in favor of informativity. *Cognition*, *209*, 104491.

Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language*, *40*(1), 136–150.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of online reading paradigms: Eye tracking, moving-window, and maze. *Journal of Psycholinguistic Research*, *41*(2), 105–128.

Zehr, J., & Schwarz, F. (2018). *PennController for internet based experiments (IBEX)*. https://doi.org/10.17605/OSF.IO/MD832