Proceedings of SALT 30: 694–713, 2020

Asymmetries between uniqueness and familiarity in the semantics of definite descriptions*

Sadhwi Srinivas
Johns Hopkins University

Kyle Rawlins
Johns Hopkins University

Daphna Heller
University of Toronto

Abstract In over a century of research into the English definite article *the*, two main theoretical factors have been identified as relevant to its meaning: namely, (i) uniqueness, and (ii) familiarity. The identification of these two factors has led to an extensive debate in semantics about which of them is more fundamental to the meaning of *the*. In this paper, we contribute to this debate by introducing novel data obtained through two controlled psycholinguistic experiments. We manipulated uniqueness and familiarity of potential referents, examining how these factors affect the comprehension and production of English definite descriptions. The behavioral results reveal an asymmetry between these two factors, with familiarity being a weaker cue than uniqueness—a pattern that is unexpected under any existing theory of definiteness. We close with a discussion of possible extensions to existing theories in light of this result, as well as avenues for future work.

Keywords: definite descriptions, uniqueness, familiarity, behavioral experiments, MTurk

1 Introduction

This paper investigates the semantics of definite descriptions in English, characterized by the presence of the definite determiner *the*: for example, *the sun*, *the cake in the fridge*, *the chocolate cake*, and so on. Here, we focus on their referential uses alone, namely those in which they are intended to identify a particular referent within the discourse context (Donnellan 1966). While it is generally agreed that such referring definite descriptions denote 'uniquely identifiable' referents (Gundel, Hedberg & Zacharski 1993), accounts of the definite article vary with regard to what conditions are sufficient for a referent to achieve uniquely identifiable status. Two

* We thank Veneeta Dayal, Florian Schwarz, the audiences and reviewers at SALT 30 and ELM 1, members of the Semantics lab at JHU, the Language Processing lab at the University of Toronto, and the Experimental Study of Meaning Lab at the University of Pennsylvania for their insightful questions and comments on this work.

main factors have been proposed in the literature, namely, (i) uniqueness and (ii) familiarity.

According to uniqueness theories of the definite article, a definite description is felicitous if and only if there is a unique entity in the discourse context that satisfies its descriptive content. This requirement is exemplified in (1) below, which is infelicitous out of the blue. Since classes typically involve more than one student, the definite description *the student in my class* describes more than one potential referent, thus failing to refer uniquely.

(1) (out of the blue) #The student in my class came to office hour today.

Familiarity theories, in contrast, do not impose a uniqueness requirement, but instead require the presence of a 'familiar' discourse referent. One common way for a referent to become familiar is through mention within the discourse. For example, in (2), native speakers of English tend to judge the anaphoric description *the pen* to be felicitous, despite the presence of more than one pen in the context.

Context: There is more than one pen on the table, only one of them is red.

(2) The red pen$_i$ that's on the table is my favorite. My grandfather gave me **the pen**$_i$ as a birthday present last year.

The question of which of these two factors, uniqueness or familiarity, is fundamental to the meaning of the definite article has received much attention within the formal semantics literature: some have advocated for uniqueness (e.g., Russell (2005), Strawson (1950), Barker (2004), Löbner (1985)), others have advocated for familiarity (e.g., Christophersen (1939), Heim (1982), Kamp (1981)), and yet others have argued that both factors are needed to account for the full distribution of the definite article (e.g., Schwarz (2009), Farkas (2002), Rawlins (2005), Roberts (2003)).

In this paper, we contribute to this discussion by way of presenting novel behavioral data from two experiments in which we manipulated uniqueness and familiarity. Given that both uniqueness and familiarity have been claimed to affect the use of definite descriptions, we believe that a quantitative evaluation of the contribution of these factors can advance our understanding of these theoretical constructs. Our main finding reveals an asymmetry between uniqueness and familiarity in both comprehension and production, where familiarity is a weaker cue than uniqueness. This is a result not obviously anticipated by any existing theory of definiteness, all of which assume that the effect of these factors is categorical.

A detailed discussion of these experiments is given in Sections 2 and 3. In the remaining part of this introductory section, we will discuss some specific construals of uniqueness and familiarity that theories have relied on, which we have drawn upon in our experimental design. In particular, the variation in how uniqueness and

familiarity have been defined can be viewed as choice points for our design.

1.1 Uniqueness theories

While all uniqueness theories require the existence of a unique object satisfying the definite description within a contextually restricted domain, one main dimension along which they vary from each other is in how exactly they delimit this domain. Standard versions of the uniqueness theory have tended to think of the referential domain as being more or less fixed. For example, Russell (2005), Evans (1977), Barker (2004) and Löbner (1985) take the domain to be delimited by the (global or immediate) deictic context described by the utterance (*semantic uniqueness*, terminology due to Roberts (2003)). A slightly weaker notion of uniqueness, namely *informational uniqueness*, is proposed in Roberts (2003), where a referent is informationally unique if and only if it is the sole entity in the common ground shared between the interlocutors to satisfy the definite description.

In this study, we examine situations that demarcate the same referential domain under both semantic and informational uniqueness. That is, in the contexts described within our experimental trials, both speaker and hearer of the definite description have equal and full knowledge of the available referents. Our experiments thus test the effectiveness of semantic/ informational uniqueness against familiarity, without distinguishing between these two types of uniqueness.

More recent uniqueness-based proposals for definiteness allow for the identity of the referential domain to be more variable. For example, Schwarz (2009) allows uniqueness to be computed in one of several possible referential domains or *situations*. Heller, Parisien & Stevenson (2016) go one step further in proposing that more than one domain is simultaneously considered (but weighted differently) in interpreting a definite description. The notion of referential domains is not directly manipulated within our experiments, but will prove relevant in considering how to interpret the observed results. We revisit this point in Section 4.

1.2 Familiarity theories

According to familiarity theories, a definite description denotes a referent that is familiar within the discourse context, regardless of whether it is unique in any strong sense. The main question that arises within these theories then is that of what makes a referent familiar. Two influential notions of familiarity exist in the literature: *strong familiarity* (Heim 1982), and *weak familiarity* (Roberts 2003). Once again, the terminology strong vs. weak is due to Roberts (2003).

The canonical way for a referent to become strongly familiar is through explicit mention of the referent in discourse, like in (2), though high degrees of salience are

also often said to contribute toward strong familiarity. For instance, Heim (1982) says that her condition of strong familiarity requires "merely that there be a unique *relevant* cat, or a unique *most salient* cat, or a unique cat that is the *most likely referent*, or something of this sort."[1,2]. In some cases where a referent is not strongly familiar, but whose identity may be easily inferred by the hearer within the discourse context (e.g., an utterance of *the sun* out of the blue), strong familiarity theories allow for a process of *accommodation*, in which a novel referent may be added to the discourse context prior to the interpretation of the definite description.

Roberts (2003) introduces the notion of weak familiarity, satisfied by any entity whose existence is entailed by the context; if an entity is strongly familiar it is weakly familiar, but not vice versa. On Roberts' view, weak familiarity is a necessary condition for the licensing of definite descriptions in addition to informational uniqueness. This is motivated by a range of commonly occurring counterexamples to strong familiarity, such as those in which speakers may refer to an entity that has not been previously mentioned but whose existence may be inferred within the discourse context—for e.g., the bridging context in (3) taken from Roberts (2003: 300).

(3) John was murdered yesterday. The knife lay nearby.

In the experiments described here, we leave open the question of how to assess weak familiarity and manipulate only *strong* familiarity: whether an entity has been explicitly mentioned. Weak familiarity, in contrast, is held constant: all referents are known to the speaker and hearer to exist in the given context.

1.3 Hybrid theories

A third class of theories, which we term here as hybrid theories, take both uniqueness and familiarity to play a role in the semantics of definite descriptions. Such hybrid theories may further be divided into two types: (i) those that regard some notion of both uniqueness and familiarity as simultaneously necessary (e.g., Roberts (2003), Rawlins (2005)), and (ii) those that regard either as independently sufficient (e.g., Schwarz (2009), Farkas (2002), Beaver & Coppock (2015)).

We have already seen one example of the first of these types, namely Roberts (2003) who takes both weak familiarity and informational uniqueness to be necessary

1 Note also the use of the word 'unique' in the quote from Heim (1982). This indicates a need for uniqueness among strongly familiar referents for successful interpretation of the description. However, under familiarity theories, this requirement is not conventionally encoded as a presupposition in the definite article: it only emerges as a downstream requirement to allow successful referent identification.

2 Other proposals have also sporadically tended to invoke the idea of maximal salience for characterizing the referent picked out by definite descriptions: e.g., Lewis (1979) and Von Heusinger (2004).

for a definite description to be licensed[3]. Among proposals of the second type, there is further variation with respect to whether both factors must be incorporated within a single, unified lexical entry for *the*, or whether there must be two separate lexical entries: one corresponding to uniqueness and the other to familiarity. While Farkas's (2002) theory of *Determined Reference* as well as the proposal in Beaver & Coppock (2015) are examples of the former, unified approach, Schwarz (2009) exemplifies the latter, ambiguity-based approach. Based on languages like German which morphologically separate uniqueness-denoting definite articles from familiarity-denoting ones, Schwarz proposes that both these factors are independently involved in the semantics of definiteness. The natural extension of Schwarz's proposal to English is to posit that there must be two independent lexical entries for *the* (though no concrete proposal has yet been made along these lines for English).

1.4 The current study

We conducted two experiments, manipulating (semantic or informational) uniqueness and strong familiarity. Specifically, we operationalized uniqueness by manipulating whether or not there was a unique intended referent in the context satisfying the descriptive content of the definite description. We operationalized strong familiarity by way of mentioning one of the referents. Our set-up allows us to make clear and concrete predictions with respect to the proposals described above. Observing successful resolution of definite descriptions only in conditions where a unique referent is present, but not in conditions with a non-unique, familiar referent, will provide support for theories relying on uniqueness. On the other hand, observing successful referential behavior in the presence of familiar but not necessarily unique referents will provide support for strong familiarity theories. Success across the board in conditions where the referent is either unique or familiar will provide support for the hybrid theories. Importantly, all predictions that can be directly derived from existing theories are categorical, and as such, observing gradient behavioral patterns is expected to pose a general challenge to any of them.

We tested the effects of these manipulations on two tasks: a comprehension experiment (Experiment 1) and a production experiment (Experiment 2).

2 Experiment 1: Comprehension

In this experiment, participants were presented with a range of stories. Each story included two referents that varied in whether they were uniquely described by the definite description contained in the final sentence, as well as whether they had been

3 However, this seems to ultimately be a redundant specification, since weak familiarity as described in Roberts (2003) is entailed by informational uniqueness.

mentioned prior to the utterance of the final sentence. Participants were asked about the interpretation of that definite description. In response, they could choose either of the referents introduced in the context, or they could answer “I don’t know." We included this latter option in order to provide the participants with a natural way to respond in cases where they were unsure as to the identity of the intended referent—especially within some of our trials which contained neither a uniquely described nor a familiar referent. In such trials where the hearer would pick the option “I don’t know”, the hearer may have responded with a clarification question within a naturally occurring, conversational context.

2.1 Method

2.1.1 Participants

We recruited thirty two participants on Mechanical Turk, all of them located within the US and self-reported native speakers of English (average age = 36 years). Each participant was compensated with $2 USD upon completing the task. This study, as well as the one reported in Section 3, were approved by the Johns Hopkins University Institutional Review Board. Participants indicated their consent for participation by clicking on an ‘Agree’ button after reading an information letter.

2.1.2 Materials

On each trial, participants read a story containing two potential referents, and interpreted a critical definite description at the end of each story. The context in each trial consisted of two interlocutors and two potential referents. For example, in the trial shown in Table 1, the two interlocutors are the chef and his assistant, and the two potential referents are the two baked goods.

Two factors were manipulated within a 2 x 2 within-subjects design: uniqueness (*-unique*, *+unique*) and familiarity (*-familiar*, *+familiar*). In the *-unique* conditions, there were two referents that fit the descriptive content of the critical definite description. For example, in the left column in Table 1, two cakes are introduced, and so the critical description *the cake* (highlighted in gray) could apply to either cake. In contrast, within the *+unique* conditions, there was only one referent that fit the descriptive content of the critical definite. For example, in the right column in Table 1, a cake and a pie are introduced, and so the critical description *the cake* (highlighted in gray) only applies to the cake and not to the pie. Next, in the *-familiar* conditions (the top row in Table 1), the intended referent is not mentioned in the dialogue between the interlocutors prior to the critical description, whereas in the *+familiar* conditions (the bottom row in Table 1), the intended referent is

	[-unique]	**[+unique]**
[-familiar]	A chef and his assistant were working together in the kitchen. On the counter, there was **a cake** containing dried fruit and **a cake** filled with jam. Then, the chef said, "Can you put **the cake** in the fridge? It must be cool before we frost it."	A chef and his assistant were working together in the kitchen. On the counter, there was **a pie** containing dried fruit and **a cake** filled with jam. Then, the chef said, "Can you put **the cake** in the fridge? It must be cool before we frost it."
[+familiar]	A chef and his assistant were working together in the kitchen. On the counter, there was **a cake** containing dried fruit and **a cake** filled with jam. The chef's assistant said, "The cake with jam looks delicious! I think we did a good job." Then, the chef said, "Can you put **the cake** in the fridge? It must be cool before we frost it."	A chef and his assistant were working together in the kitchen. On the counter, there was **a pie** containing dried fruit and **a cake** filled with jam. The chef's assistant said, "The cake with jam looks delicious! I think we did a good job." Then, the chef said, "Can you put **the cake** in the fridge? It must be cool before we frost it."

Table 1 Example stimuli showing 2x2 manipulations of uniqueness and familiarity (through mention) of referents within Experiment 1.

mentioned in the dialogue, thus making it strongly familiar.

We created thirty two stories in total. The materials were further counter-balanced across items in various aspects. For example, in the *+familiar* conditions, half of the items contained turn-taking between the interlocutors, whereas in the other half the same interlocutor acted as the speaker throughout the dialogue phase. The items were also counter-balanced in whether the intended referent was introduced as the first or the second potential referent during the set-up phase. To make the dialogues sound natural, the descriptive content of the referring expressions changed from a longer, more modified first mentions to shorter, less modified critical descriptions in the *+familiar* conditions, consistent with studies that establish this as the natural progression over time for descriptions to the same object (Clark & Wilkes-Gibbs 1986).

Each story was paired with a comprehension question regarding the identity of the referent intended by the speaker of the critical definite description. For instance, for the trial shown in Table 1, they were asked: *"Which object does the chef want his assistant to put in the fridge?"*. In response, participants could choose one of three possible responses: the intended referent (e.g., the cake filled with jam), the other referent (e.g., the pie/cake containing dried fruits), or the option "I don't know". The order of the two referents was counter-balanced across participants: while for half

of the participants, the intended referent appeared first among the three options, it appeared second for the other half. "I don't know" always appeared as the third and final option. Each story was instantiated in all four conditions: [*-unique, -familiar*], [*-unique,+familiar*], [*+unique, -familiar*], and [*+unique, +familiar*]. However, no participant saw the same story in more than one condition. Instead, each participant saw eight stories in each of the four conditions, namely thirty two trials in total.

2.1.3 Procedure

Participants performed the task online through Amazon's Mechanical Turk interface. Participants were told that they would read short stories and answer questions about them. They were instructed to read the stories carefully, and answer the question that follows. We told participants that in some cases the answer may be obvious, while in other cases, they may be less sure. Participants saw each story on a new screen: they could not skip any trial, and could not go back and change their answers. All in all, the task lasted about 20 minutes on average.

2.2 Results

Figure 1 depicts the full pattern of choices made by participants within each of the four conditions. Each bar is divided into regions of varying darkness, with each region representing the proportion of trials where participants chose a particlar type of referent. The darkest regions denote the proportion of choices to the intended referent (that is, the uniquely described referent in the *+unique* conditions, and the familiar referent in [*-unique,+familiar*]), the lighter regions denote the proportion of choices to the competing referent, and the lightest regions denote the proportion where participants chose the option "I don't know".

We examined these patterns statistically by fitting a mixed-effects logistic regression model with repeated contrasts (Schad, Vasishth, Hohenstein & Kliegl 2020). In effect, this method treats the combination of uniqueness and familiarity as a single, four-level manipulation This coding scheme allowed us to make the following three comparisons: (i) comparing [*-unique,-familiar*] to [*-unique,+familiar*] to test the effect of familiarity, (ii) comparing [*-unique,+familiar*] to [*+unique,-familiar*] to compare familiarity alone against uniqueness alone, and (iii) comparing [*+unique,-familiar*] to [*+unique,+familiar*] to ask whether familiarity has any effect when uniqueness is already satisfied. The dependent binary variable indicated whether the intended referent had been correctly chosen within a trial (coded as 1), or not (coded as 0): this corresponds to the darkest regions of the bars in Figure 1. The model also included random intercepts and all random slopes for participants and items.

We begin our discussion of the results by first considering participants' behavior

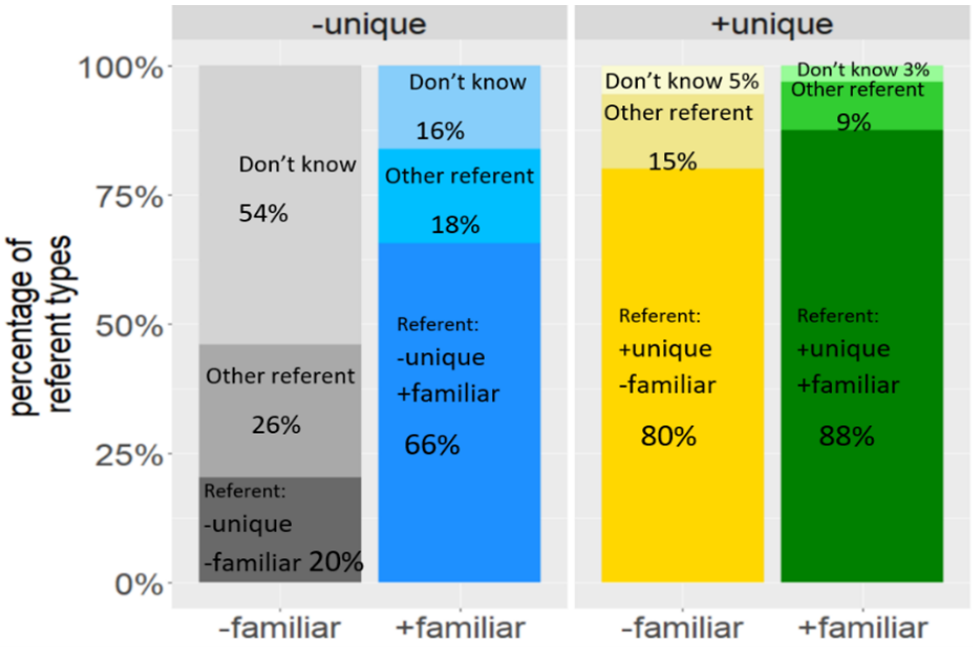

Figure 1 The referent participants chose across the four conditions in response to the critical definite description: the intended referent (darkest bars), the other referent (lighter bars), or "I don't know" (lightest bars).

in the [*-unique,-familiar*] condition (grey bar in Figure 1). As expected, participants were the least sure in this condition, choosing the option "I don't know" 54% of the time, and the two possible referents roughly equally: 20% for the intended referent, and 26% for the other referent. In the [*-unique,+familiar*] condition (blue bar), the intended referent was now chosen more often at 66%. Our model indicated that this was a significant increase from ($\beta = 2.6$, $SE = 0.43$, $z = 6.1$, $p < .001$), indicating that making a referent familiar by mention makes it a more likely candidate to be picked out by a definite description. Next, when the intended referent was instead unique but not familiar [*+unique,-familiar*] (yellow bar), it was selected at 80%: importantly, this is a significant increase from the [*-unique,+familiar*] condition ($\beta = 1.3$, $SE = 0.38$, $z = 3.4$, $p < .001$). This difference between between uniqueness and familiarity is surprising under any theory that takes them to be equal cues. Finally, in the [*+unique,+familiar*] condition, the intended referent was chosen numerically more at 88% compared to the [*+unique,-familiar*] condition; however, this difference was not significant ($\beta = 0.5$, $SE = 0.42$, $z = 1.25$, $p = .21$), indicating that familiarity and uniqueness do not together provide a stronger cue than uniqueness alone.

2.3 Discussion

In line with existing theories of definiteness, our experimental results show that each of uniqueness and familiarity (instantiated in the form of previous mention) can individually influence the interpretation of a definite description. However, our findings contribute a novel perspective as they reveal that uniqueness is a stronger cue than familiarity, an asymmetry that is unexpected under existing theories.

One important discrepancy between the predictions made by existing theories and what we have observed in our experiment pertains to the expectations within these theories of categorical patterns of behavior. For instance, standard uniqueness theories relying on the notions of semantic or informational uniqueness tend to predict complete failure of reference in the absence of a uniquely-described referent, without regard to whether one of the non-unique, potential referents is familiar. Similarly, standard familiarity theories assume complete referential success in the presence of a familiar referent, even if a stronger notion of uniqueness is violated, and thus do not anticipate any difference in behavior between [-*unique,*+*familiar*] and the [+*unique,*+*familiar*] conditions. Categorical predictions are made by the hybrid theories as well: while these do expect effects of both uniqueness and familiarity, they are taken to be equally necessary or independently sufficient cues, and as such, no asymmetry is predicted. However, our experimental results point towards the need for a theory of definite description comprehension that anticipates systematic gradience of the form observed. Furthermore, the fact that we found that both uniqueness and familiarity serve as a cue for referent choice, suggests that some version of a hybrid theory is needed.

However, before we turn to consider how to extend existing theories in light of the observed result, we need to consider a possible objection to our experimental task. Specifically, one may argue that the observed pattern of comprehension alone may not reflect the felicity of uniqueness and familiarity as cues in the interpretation of definite descriptions. In other words, while participants may be less likely to choose the intended referent in [-*unique,*+*familiar*], they nevertheless find the definite description as felicitous in this context as in the +*unique* conditions. We argue that our results show that both uniqueness and familiarity lead to the felicity of a definite description: either cue alone leads to choosing the intended referent on the majority of trials. However, we further argue that considering quantitative measurements of the preferences of interpretation can advance our understanding of these cues. Nevertheless, to ensure that the asymmetry between uniqueness and familiarity is not a by-product of the forced choice task, we conducted a second experiment where referring expressions were not interpreted by rather produced.

A different concern with the current setup is the operationalization of familiarity. Recall that in each trial, the context introduced the interlocutors and potential

referents which created the set up for the dialogue. Throughout in our discussion, we have assumed that a referent becomes strongly familiar only if it has been mentioned prior to the critical description in the dialogue phase of the story. However, we note that the contextual set-up was also done via the use of language, and so the initial introduction of the referents may have sufficed to make them 'familiar' from the perspective of the participants, even though the referents were not strongly familiar to the interlocutors in the story. To address this issue, we are planning to introduce the referents non-linguistically in further iterations of this study, through the use of images[4]. However, within the current study, note that even if both potential referents were strongly familiar in our stories, the observed results nevertheless reveal an asymmetry between uniqueness and familiarity that is not predicted by any existing theory. The presence of more than one familiar referent is expected to lead to failure of interpretation even under strong familiarity theories like Heim's as well as weak familiarity theories like the one in Roberts (2003), which do not anticipate the partial success that we observed in identifying the target in [*-unique,+familiar*].

3 Experiment 2: Production

In this experiment, participants were presented with the same context stories as in Experiment 1, except the final definite description was replaced by an empty text box, and participants were asked to provide a natural completion to the story.

3.1 Method

3.1.1 Participants

Thirty two participants in the United States, all of whom were self reported native speakers of English, were recruited via Mechanical Turk (average reported age = 38 years). Each of them was compensated with $2 USD for their time.

3.1.2 Materials

The materials were identical to those used in Experiment 1, except that the critical definite description was now replaced with an empty text box—see Table 2.

4 Some readers may object to this as well: even if the introduction of the referent is not via explicit mention, it may still be 'salient' enough to be strongly familiar. We acknowledge this concern, and further note that it is tricky to address what exactly constitutes 'salience' of a referent, and how much salience is enough to constitute strong familiarity. Without a precise theory that does not simply rely on an intuitive notion of this concept, it is difficult to control this factor within an experiment such as the one described here. But one solution may be to view familiarity or salience as a gradient concept to begin with. We revisit this point in the general discussion of our findings in Section 4.

	[-unique]	[+unique]
[-familiar]	A chef and his assistant were working together in the kitchen. On the counter, there was **a cake** containing dried fruit and **a cake** filled with jam. Then, the chef said, "Can you put [redacted] in the fridge? It must be cool before we frost it."	A chef and his assistant were working together in the kitchen. On the counter, there was **a pie** containing dried fruit and **a cake** filled with jam. Then, the chef said, "Can you put [redacted] in the fridge? It must be cool before we frost it."
[+familiar]	A chef and his assistant were working together in the kitchen. On the counter, there was **a cake** containing dried fruit and **a cake** filled with jam. The chef's assistant said, "The cake with jam looks delicious! I think we did a good job." Then, the chef said, "Can you put [redacted] in the fridge? It must be cool before we frost it."	A chef and his assistant were working together in the kitchen. On the counter, there was **a pie** containing dried fruit and **a cake** filled with jam. The chef's assistant said, "The cake with jam looks delicious! I think we did a good job." Then, the chef said, "Can you put [redacted] in the fridge? It must be cool before we frost it."

Table 2 Example of stimuli used in Experiment 2. The grey box represents the text box where participants freely typed in their response.

3.1.3 Procedure

Participants performed the task online through Mechanical Turk. Participants were told that they would read stories in which some details are missing, and their task was to fill in the missing information. Participants were further informed the missing information may be more obvious in some case and less obvious in other cases. Importantly, we did not give participant any further instructions about what information they should put in the text box.

3.2 Results

While the task was highly unconstrained, the context restricted grammatical completions to noun phrases. Indeed, participants produced appropriate noun phrases on 80% of the trials. The most common noun phrase type was definite descriptions (55%), but participants also produced demonstratives (14%), pronouns (6%), and even indefinite descriptions (5%). In a few cases, participants produced bare nouns, which are not grammatical.

We focus our attention on the produced definite descriptions, specifically those descriptions that refer to the intended target referent. Such descriptions amounted to

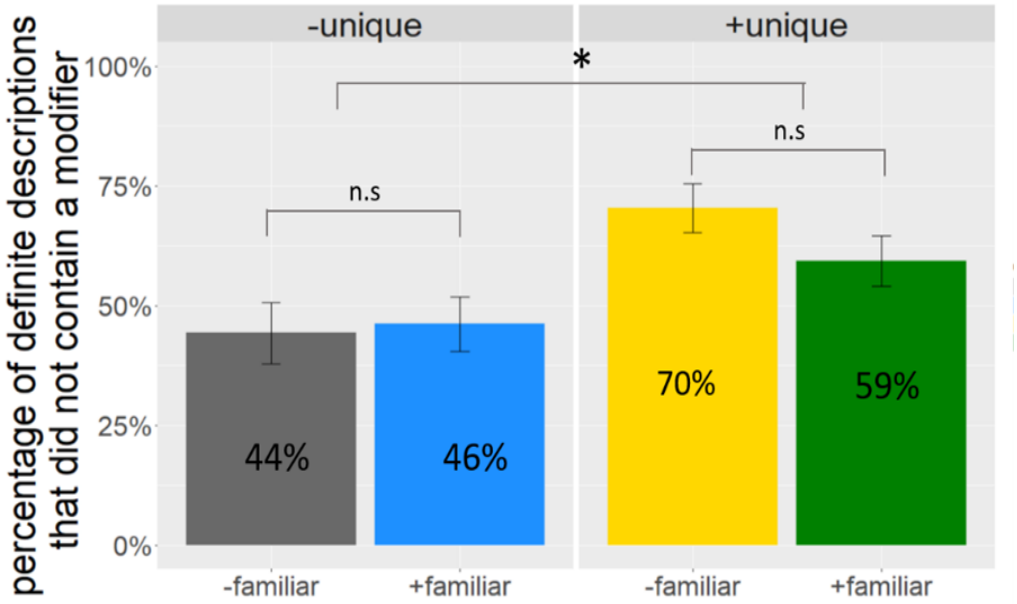

Figure 2 % of bare definite descriptions (without any disambiguating modifier) used across the four conditions of uniqueness and familiarity.

about 30% of the observed data (though definite descriptions amounted to 55% of the observed data, about half of these clearly referred to something other than the target, such as the distractor: *the pie with dried fruit*, the target and the distractor: *the cakes*, or an altogether different object in the scene). In particular, we examine the proportion of definite descriptions that were produced without any adjectival or prepositional modifiers, namely expressions of the form *the cake* for the example shown in Table 2, instead of expressions like *the cake with jam* or *the jam cake*. Our logic is that if participants deemed the preceding context to license an unmodified definite description for the intended referent, they would be more likely to use the unmodified description (*the cake*) over a longer, presumably costlier and more redundant modified version (*the cake with jam*). Thus, in the *+unique* conditions in Table 2, if the presence of a unique cake is sufficient to license the description *the cake*, then people would likely use this description over *the cake with jam*. Similarly, in the *+familiar* conditions, if familiarity (through previous mention) of the cake filled with jam is sufficient to license the unmodified definite description *the cake*, people are expected to use an unmodified description.

Figure 2 plots the proportion of unmodified definite descriptions across the four conditions. We examined these data statistically by fitting a 2x2 mixed-effects logistic regression model. The dependent variable was whether the definite descrip-

tion produced to describe the intended referent included a modifier (coded as 0) or not (coded as 1). The two categorical independent variables were coded using weighted-effects coding (using the **wec** package in R; Te Grotenhuis, Pelzer, Eisinga, Nieuwenhuis, Schmidt-Catran & Konig (2017)), which has important advantages over traditional effect coding when analyzing unbalanced data such as in our case, where restricting the analysis to a subset of the production data resulted in different numbers of observations for each level of uniqueness and familiarity. This resulted in the following contrasts: (a) uniqueness: *-unique* -1.2 vs. *+unique* 1; (b) familiarity: *-familiar* -1.15 vs. *+familiar* 1. In choosing the random effects, we considered all random effect structures ranging from simplest to most complex, selecting the one with the best AIC fit. The best-fitting model included a random intercept for items, a random intercept for participants, and a random uniqueness slope for participants.

Regression results revealed that unmodified definite descriptions were produced significantly more in the *+unique* than the *-unique* conditions (65% *vs.* 45% in Figure 2: β = 1.1, *SE* = 0.35, z = 3.1, p = .002), indicating that uniqueness in the discourse context licensed more bare descriptions, compared to context which included two objects of the same kind. In contrast, familiarity did not have the same effect: a [*-unique, +familiar*] referent was referred to with an unmodified definite just as much as a [*-unique, -familiar*] referent (46% vs. 44%: β = -0.27, *SE* = 0.18, z = -1.5, p = .14). This indicates that familiarity alone does not render the referent appropriate for an unmodified definite description in a parallel way to uniqueness, thus mirroring the asymmetry between uniqueness and familiarity observed in Experiment 1.

3.3 Discussion

The results of Experiment 2 also revealed a similar asymmetry as in Experiment 1, since the presence of a uniquely described referent led to the use of more unmodified subsequent descriptions to the target referent than the presence of a familiar (mentioned) but NOT uniquely described referent. Taken together, the results from both experiments constitute novel evidence that both uniqueness and familiarity of a referent can license definite descriptions in English, albeit to varying degrees.

We might wonder why the proportion of unmodified descriptions was not different between [*-unique,-familiar*] and [*-unique,+familiar*], contrasting with the comprehension results that showed a significant effect of familiarity. We speculate that this could arise from an asymmetry between comprehension and production, wherein the cooperative hearer maximizes the utilization of the cues available in interpretation, while the cooperative speaker prioritizes the use of stronger cues over weaker ones. An alternative hypothesis emerges if, following Kehler & Rohde (2013), any factor that helps with comprehension alone but not production is one that

is not directly relevant to the lexical semantics of the referring expression. Under this view, we might then expect that uniqueness is lexically encoded in the definite article, but familiarity effects follow from more general pragmatic principles that influence comprehension of reference. This would provide an alternative view of the observed uniqueness-familiarity asymmetry which supports a categorical uniqueness theory of the definite article over the hybrid one. This interpretation is further discussed in Section 4 below.

We might additionally wonder about is why participants did not produce more pronouns in the *+familiar* conditions, where they arguably sound more natural than even unmodified definite descriptions. Here, it may have been the case that participants' exposure to trials belonging to the *-familiar* conditions which did not license the use of pronouns led them to disprefer the use of pronouns in *+familiar* as well (a task effect, akin to 'lexical entrainment', cf. Garrod & Anderson (1987)).

4 General discussion

In this paper, we described two behavioral experiment that examine the comprehension (Experiment 1) and production (Experiment 2) of English definite descriptions, examining the effects of uniqueness and familiarity on the quantitative patterns of reference comprehension and production. Both comprehension and production patterns revealed an asymmetry between uniqueness and familiarity. In comprehension, both uniqueness and familiarity independently licensed a definite description, but familiarity was a weaker cue. In production, familiarity led to fewer unmodified definite descriptions.

What does the observed asymmetry mean for existing theories of definiteness? Given that we found an effect of both uniqueness and familiarity in comprehension, some version of a hybrid theory seems necessary at a first glance to explain the full distribution of the definite article. However, no existing hybrid theory is appropriate, since (i) they predict only categorical results, and (ii) they predict uniqueness and familiarity to be equally effective, and thus do not anticipate the observed asymmetry between them. While it may be possible to probabilistically extend existing proposals, there is more than one way to do so in light of the experimental results. Here, we discuss two options: one that incorporates probabilistic uncertainty in computing a referent's uniqueness, and another that incorporates a probabilistic notion of familiarity.

While uniqueness of a referent relative to a definite description is generally categorical within a given referential domain (the referent is either the only item in the domain that the description can apply to, or it is not), uncertainty about a referent's uniqueness may be indirectly introduced via uncertainty about the referential domain itself. The idea of uncertainty in referential domains has been most explic-

itly discussed within the psycholinguistic literature on the processing of referring expressions, particularly in cases involving a mismatch between the knowledge state of the hearer and the speaker (Heller et al. (2016), Mozuraitis, Stevenson & Heller (2018)). These studies show that interlocutors' observed behavior in interpreting definite descriptions in such situations is most accurately explained by assuming that they simultaneously consider both their own perspective and their partner's. Rephrased in terms of referential domains, this means that the privileged interlocutor computes reference simultaneously in two domains: one corresponding to their own knowledge state which does not exclude the privileged object, and another corresponding to their conversational partner's which does exclude this item.

This idea of simultaneous consideration of more than one referential domain can be extended to our experimental trials as well. Specifically, we might take the deictic context surrounding the dialogue (i.e, what is introduced in the set-up phase of our stories) as one candidate for the referential domain—let us call this domain RD-1, while the narrow discourse context consisting solely of referents that have been explicitly mentioned is another candidate (RD-2). Thus, in the [-*unique,*+*familiar*] condition in Table 1, RD-1 consists of both the cakes but RD-2 only consists of the previously mentioned jam-filled cake. The intermediate proportion of choices to the familiar referent observed within this condition suggests that a weighted average of both these domains is considered. On the other hand, in [+*unique,*+*familiar*], the intended referent is uniquely described in both domains, and will therefore be picked invariably of the weights associated with each domain. To decide whether this alternative is indeed viable, we need a better understanding of the types of referential domains that interlocutors are sensitive to: for instance, whether domains are restricted to situations associated with a place, time and point of view, or can they be defined more narrowly to consist only of some parts of a discourse context but not others. We would also need to test the comprehension and production of definite descriptions within experimental stimuli that explicitly manipulate this notion.

A second alternative would be to introduce uncertainty about the familiarity status of the intended referent. As stated above, under a particular choice of referential domain, uniqueness of the referent given a description is categorical. However, familiarity may be graded, wherein entities may be more or less familiar. Under this view, previous mention of a potential referent in our experiments would be viewed as increasing that referent's familiarity by a significant—but crucially, not maximal—level, resulting in the graded result in the [-*unique,*+*familiar*] condition. This option too requires further empirical testing to attest viability. For instance, follow-up norming studies would need to independently measure whether familiarity is indeed graded, the factors that lead to an increase or reduction of the familiarity of a referent, and whether success in comprehension experiments is correlated with independent measurements of familiarity—regardless of the referential domains.

Apart from choosing how to introduce uncertainty within the model, attempting to probabilistically extend existing theories in the ways discussed above leaves us with yet another dimension of choice: namely, whether to posit a unified lexical entry for the definite article, or whether to posit ambiguity. While a unified semantics may be somewhat more parsimonious for English (which does not morphologically differentiate between the uniqueness and familiarity articles), the fact that there exist languages that separate the two forms, along with a universality assumption that all languages of the world instantiate lexical items with identical underlying semantics might bias us towards positing ambiguous semantics for English *the* as well. Both unified and ambiguity-based analyses are thus possible in principle, but we would like to point out that the graded results we observed in our experiments speak in favor of a unified account, at least at a first glance. This is because within an ambiguity analysis, making only one of the lexical entries encode probabilistic gradience (e.g., the one corresponding to familiarity) makes the other, categorical entry (e.g., the one corresponding to uniqueness) redundant—since in the absence of further stipulations, the probabilistic entry is powerful enough to also handle the categorical edge case (e.g., a case where a referent is uniquely described though not maximally familiar).

Finally, a third possibility is to take as our starting point the observation that participants were NOT more likely to describe the familiar referent in [*-unique,+familiar*] with an unmodified definite description than the non-unique, non-familiar referent in [*-unique,-familiar*]. This opens up yet another interpretation of the observed results favoring a categorical uniqueness theory of definiteness. Such an interpretation crucially involves adopting Kehler & Rohde's (2013, 2019) proposal that comprehension and production processes in reference do not exactly mirror each other. According to this, comprehension of referring expressions is affected by general considerations of discourse coherence in addition to the lexical meaning encoded within the expression, but their production is only affected by the latter[5]. In other words, the meaning encoded in the referring expression (definite descriptions, in our case) is reflected most clearly in production but not comprehension. For our experiments, this would mean that uniqueness alone is lexically encoded in the definite article, while familiarity of the referent only serves to affect discourse

5 More precisely, Kehler & Rohde (2013, 2019) propose that comprehension and production are systematically related to each other via the Bayes' rule, as shown below:

(i) $P(\text{referent} \mid \text{expression}) \propto P(\text{expression} \mid \text{referent}) * P(\text{referent})$

According to this expression, the probability of interpreting a referring expression to denote a particular referent $P(\text{referent} \mid \text{expression})$ depends on whether the lexical meaning of the expression makes it appropriate to describe that referent $P(\text{expression} \mid \text{referent})$, as well as the prior probability $P(\text{referent})$ of the referent being mentioned at all, as determined by coherence considerations.

coherence expectations. Note that this explanation requires us to adopt a more unconventional processing model than the one assumed to be default in the literature, wherein comprehension and production processes necessarily mirror one another (see Kehler & Rohde (2019) for further discussion of this point). Testing this alternative would require independently measuring whether familiarity does indeed affect coherence expectations, gauging the extent to which speakers use unmodified definite descriptions for a given referent using sentence completion tasks.

More broadly, the current study illustrates the importance of quantitatively evaluating semantic theories by examining fine-grained patterns of language comprehension and production that may reveal gradient patterns that are not easily detectable using a small number of examples. The preceding discussion also brings out the necessity for simultaneous evaluation of semantic competence accounts (usually investigated within formal semantics via the use of introspective examples) and performance or processing accounts (investigated within psycholinguistics through controlled experiments), since the competence account that one chooses implicitly fixes a processing account when explaining behavioral results, and vice versa.

5 Conclusion

In sum, our experimental investigation of uniqueness and familiarity in the comprehension and prediction of definite descriptions revealed a novel asymmetry between these two factors, where familiarity is a weaker cue than uniqueness. This finding cannot be accounted for without extending existing analyses of definiteness in the theoretical literature, all of which predict categorical judgments rather than systematic gradience. Many existing proposals are compatible with several probabilistic extensions that could potentially account for the data we present here, but further empirical work would be necessary to decide between the available options.

References

Barker, Chris. 2004. Possessive weak definites. *University of Massachusetts Occasional Papers* (29). 89.

Beaver, David & Elizabeth Coppock. 2015. Novelty and familiarity for free. In Thomas Brochhagen, Floris Roelofsen & Nadine Theiler (eds.), *20th Amsterdam Colloquium*, 50–59.

Christophersen, Paul. 1939. *The Articles: A Study of their Theory and Use in English*. Einar Munksgaard.

Clark, Herbert H & Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition* 22(1). 1–39.

Donnellan, Keith S. 1966. Reference and definite descriptions. *The philosophical review* 75(3). 281–304.

Evans, Gareth. 1977. Pronouns, quantifiers, and relative clauses (I). *Canadian journal of Philosophy* 7(3). 467–536.

Farkas, Donka F. 2002. Specificity distinctions. *Journal of semantics* 19(3). 213–243.

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

Gundel, Jeanette K, Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 274–307.

Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*: University of Massachusetts, Amherst PhD dissertation.

Heller, Daphna, Christopher Parisien & Suzanne Stevenson. 2016. Perspective-taking behavior as the probabilistic weighing of multiple domains. *Cognition* 149. 104 – 120. doi:https://doi.org/10.1016/j.cognition.2015.12.008. http://www.sciencedirect.com/science/article/pii/S0010027715301244.

Kamp, Hans. 1981. A theory of truth and semantic representation. *Formal semantics-the essential readings* 189–222.

Kehler, Andrew & Hannah Rohde. 2013. A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics* 39(1-2). 1–37.

Kehler, Andrew & Hannah Rohde. 2019. Prominence and coherence in a Bayesian theory of pronoun interpretation. *Journal of Pragmatics* 154. 63–78.

Lewis, David. 1979. Scorekeeping in a language game. In *Semantics from different points of view*, 172–187. Springer.

Löbner, Sebastian. 1985. Definites. *Journal of semantics* 4(4). 279–326.

Mozuraitis, Mindaugas, Suzanne Stevenson & Daphna Heller. 2018. Modeling reference production as the probabilistic combination of multiple perspectives. *Cognitive science* 42. 974–1008.

Rawlins, Kyle. 2005. Possessive definites and the definite article .

Roberts, Craige. 2003. Uniqueness in definite noun phrases. *Linguistics and Philosophy* 26(3). 287–350.

Russell, Bertrand. 2005. On denoting. *Mind* 114(456). 873–887.

Schad, Daniel J, Shravan Vasishth, Sven Hohenstein & Reinhold Kliegl. 2020. How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language* 110. 104038.

Schwarz, Florian. 2009. *Two types of definites in natural language*: University of Massachusetts, Amherst PhD dissertation.

Strawson, Peter F. 1950. On referring. *Mind* 59(235). 320–344.

Te Grotenhuis, Manfred, Ben Pelzer, Rob Eisinga, Rense Nieuwenhuis, Alexander Schmidt-Catran & Ruben Konig. 2017. When size matters: advantages of

weighted effect coding in observational studies. *International Journal of Public Health* 62(1). 163–167.

Von Heusinger, Klaus. 2004. Choice functions and the anaphoric semantics of definite NPs. *Research on language and Computation* 2(3). 309–329.

Sadhwi Srinivas
sadhwi@jhu.edu

Kyle Rawlins
kgr@jhu.edu

Daphna Heller
daphna.heller@utoronto.ca