

Maxim of quantity and presupposition in understanding object labels

Frisson, Steven; Murphy, Gregory

DOI:

[10.1080/23273798.2019.1650193](https://doi.org/10.1080/23273798.2019.1650193)

License:

Other (please specify with Rights Statement)

Document Version

Peer reviewed version

Citation for published version (Harvard):

Frisson, S & Murphy, G 2019, 'Maxim of quantity and presupposition in understanding object labels', *Language, Cognition and Neuroscience*. <https://doi.org/10.1080/23273798.2019.1650193>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

This is an Accepted Manuscript of an article published by Taylor & Francis in *Language Cognition and Neuroscience* on 02/08/2019, available online: <http://www.tandfonline.com/10.1080/23273798.2019.1650193>

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Maxim of Quantity and Presupposition in Understanding Object Labels

Steven Frisson

University of Birmingham

Gregory L. Murphy

New York University

Please address all correspondence to:

Steven Frisson

School of Psychology

University of Birmingham

52 Pritchatts Road

Edgbaston

B15 2TT

UK

Email: s.frisson@bham.ac.uk

Abstract

Two experiments investigated whether listeners mandatorily apply the Gricean maxim of quantity. In conversation, it is infelicitous to use a label that does not distinguish the referent (“the car” when there are two cars) or to use an unnecessarily specific label (“the convertible” when only one car is present). Subjects verified labels that picked out one of two objects, which was visually cued. Labels were at the basic or subordinate level, and the objects were related (two cars) or not (car and boat). Presuppositions were manipulated by varying determiners (“the/a car”). Responses were slower and less accurate when the basic-level name was used in the context of two related objects, suggesting that listeners prefer more distinguishable names even the referent is unambiguous. Overly specific names did not incur a cost. Manipulating determiners had little effect. These results argue against a general use of Gricean maxims or presuppositions when identifying referents.

It has long been known that people prefer to label objects at the *basic level* of categorization, in free naming of pictures (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), in conversation (both for children and adults; Anglin, 1977; Brown, 1958; Callanan, 1990), and in written text (Wisniewski & Murphy, 1989). However, a more specific or *subordinate*-level name can be called for when finer distinctions are required (Cruse, 1977). If you say, “Put the box in the car parked outside,” there had better be only one car parked outside, unless your addressee already knows which car you are referring to. If there are multiple cars, then the instruction should be something like, “Put the box in the red Prius outside.” We will present data from two experiments that examined whether under- and over-specification of an object’s description invokes a processing cost.

The appropriate level of a description can be determined within a Gricean framework. According to the Gricean maxim of quantity (Grice, 1975), sometimes called the maxim of informativeness, one expects speakers to provide a specific enough description to distinguish an intended referent from other objects in the context. Deviating from this expectation, for example when someone uses a more specific name than would be required in the context, can result in communicative breakdown (even in preschoolers: Eskritt, Whalen, & Lee, 2008; Okanda, Asada, Moriguchi, & Itakura, 2015) or lead the hearer to assume something more was meant with the statement. For example, if Lewis says, “My Jaguar is parked outside,” one might well think that he is making the point that he owns an expensive sports car rather than neutrally providing information (Clark & Murphy, 1982; Cruse, 1977). When there is no expectation for Lewis to provide information about his car’s status, it could take a listener more time to process this reference (compared to “My car”), as such inferences often take time to draw (e.g., Haviland & Clark, 1974). However, as Cruse (1977) first noted, it is not considered infelicitous for someone

to use a basic level name even if it provides unnecessary information, e.g., referring to your family's only pet as "the dog" even though everyone knows it is a dog, and so a more general term such as "the animal" would have been sufficient to pick it out.

Outside of conversations, it is not clear whether context has a strong effect on which name is used. Rosch et al. (1976, Experiment 10) found little effect of picture set on people's labeling of pictures. For example, they tended to call chairs *chair* even when the picture set had multiple chairs and furniture, rather than giving a name that would distinguish each item. Lin, Murphy, and Shoben (1997, Experiment 4) found very little effect of prior task context on the labeling of items. Even when asked to distinguish different kinds of dogs in an initial task, people did not then change their labeling of dogs to more specific names. These results seem inconsistent with the notion that names should be specific enough to distinguish referents in context, suggesting instead that basic-level names are strongly preferred in spite of pragmatic constraints. However, these experiments focused on production measures. It is possible that listeners would find it more difficult to understand names that violated Gricean maxims. For example, referring to a picture of a car as *a convertible* might seem strange if there is no other car to distinguish it from. Even if a participant knows that the object is a *convertible*, processing the reference might take longer than when a less specific reference is sufficient and, according to the quantity maxim, preferable. Of course, subordinate labels generally do take longer to process than basic-level names (Murphy & Brownell, 1985; Rosch et al., 1976). Our point is that calling a car a *convertible* should be harder to understand when there is only one car in the context than when the *convertible* is to be distinguished from other cars.

However, it is less clear whether the same consideration applies at the basic level. That is, would referring to a depicted car as *a car* be difficult to process when another car is present? On

the one hand, the name is not distinctive and, hence, it can be considered underinformative; on the other, *car* is the neutral name for the object. If the reference is not ambiguous (e.g., the speaker is pointing at the referent or the sentence disambiguates it), as is the case in the present experiments, the basic name may seem fine even though it is not distinctive.

Do such Gricean processes take place regardless of context, or do they only operate in conversational situations where speakers' intentions are salient? It could be that listeners are sensitive to the quantity maxim only when there is a real conversation, or at least a plausible narrative. Merely labeling a picture may not evoke pragmatic factors, as in the Lin et al. (1997) study. Alternatively, it is possible that such pragmatic variables are automatically used even in situations where there is no interlocutor, given that objects are mostly labeled in actual conversations. That is, most people have little practice in understanding object labels outside of conversations and so might find it difficult to ignore the factors that normally influence labeling.

Engelhardt, Bailey, and Ferreira (2006) investigated the specificity of descriptions in both naming and comprehension in a visual world paradigm. However, they did not look at category names but instead considered whether nouns had pre- or post-modifiers that provided additional information, such as "the apple on the towel." They found that speakers often provided additional description that was not strictly necessary. Furthermore, listeners judging the quality of an instruction did not like under-descriptions, which resulted in ambiguity, but considered overdescriptions to be as good as minimal descriptions when there was a single potential referent. However, in a visual world task, eye movement patterns did indicate that overdescriptions carried a cost. Hence, as there seems to be a contradiction in the on-line and off-line results, it is unclear whether overly specific names like *convertible* in our experiment will cause listeners any trouble.

The present experiments investigated whether such Gricean factors operate in a simple name verification task, with names that were or were not pragmatically felicitous. Unlike most previous studies, which tested production, the present experiments examined comprehension. We also investigated a related pragmatic factor involving the presuppositions of descriptions, namely the presence of a definite or indefinite determiner. Definite determiners are generally used when “referring to something which can be identified uniquely in the contextual or general knowledge shared by speaker and hearer” (Quirk, Greenbaum, Leech, & Svartvik, 1985, p. 265). As a result, there is a presupposition when saying “pick the car” that there is only one car in the context or else one car that is readily distinguished from the others. Thus, if the display contains two normal cars, using the basic level noun with the definite determiner seems infelicitous. “Pick a car” would be more appropriate than “Pick the car.” In contrast, with the subordinate label, the definite determiner, such as “pick the convertible,” is appropriate in this situation, because there would only be one convertible, and so it is uniquely identifiable. In contrast, “pick a convertible” would be strange under the assumption that the convertible is uniquely identifiable and so should have received the definite determiner. It is a general pragmatic rule that speakers should choose the strongest term available in such cases (e.g., should say “the” instead of “a,” if possible, just as they should say “all” instead of “some,” even when both are true), as discussed by Schlenker (2012). Schlenker argues that this rule is also derived from the Gricean maxim of quantity. This is a technical issue that we do not need to take a position on, but for the sake of exposition, we will refer to the effects investigated here as “Gricean.”

We constructed two tasks in which subjects heard a sentence referring to one of two depicted items and then judged whether it was true or false. We labeled the object by a pragmatically felicitous or infelicitous name and compared how long it took people to verify that the test object

was correctly named. If Gricean processes are mandatory, then we should find an effect of name appropriateness even though there is no issue of referential ambiguity. That is, it should be strange to refer to one car as *the car* when there is another car in the display, even if it is unambiguous as to which car is being referred to. However, because basic-level names are generally appropriate (Cruse, 1977), it may be that they are not affected by such contextual variables. Correspondingly, it should be strange to refer to a convertible as *a convertible* when there is no other car present. Because subordinate names are not generally preferred, it may be possible that pragmatic factors have a greater influence on their comprehension.

If hearers are sensitive not only to the issue of informativeness but also to the presuppositions expressed by the determiner, there should be an interaction of determiner and category level, with the subordinate level preferring the definite determiner, but the basic level preferring the indefinite. Since psycholinguistic experiments have often found that readers are sensitive to the presuppositions of determiners (e.g., Haviland & Clark, 1974; Murphy, 1984), it seems likely that we will find such effects if our participants also follow the Gricean rule of descriptions.

Experiment 1

We presented two pictures side by side (see examples 1-4, below), either depicting instances of the same category (“related”) or a different category. A Gricean account would predict that a subordinate name would be a better description when the pictures are related and a basic-level name when they are unrelated, as the reverse combination would result in under- and over-specification. Concretely, one would expect that a basic-level term will be more felicitous when there are two pictures of different categories compared to when there the two pictures are of the same category. Finally, if people are sensitive to presupposition, then the use of a definite and indefinite determiner will affect the level of appropriateness of the name, with the definite

determiner a better fit when the pictures are unrelated and a basic-level term is used, and when the pictures are related and a subordinate term is used.

Method

Subjects. A total of 56 undergraduate students (49 female) from the University of Birmingham, UK, participated in the experiment. One subject was deleted as she did not respond in time in 81% of the cases and another because she made over 20% errors.

Materials. We created a total of 24 target items, each of which appeared in 4 conditions (see Appendix). We used colored photographs extracted from internet sources depicting familiar items from domains such as foods, furniture, clothes, animals, vehicles, and tools. Each item consisted of two pictures. In the related conditions, the two pictures depicted the same basic-level category (e.g., two cars) but two different subordinate categories (e.g., a convertible and a limo). In the unrelated conditions, the pictures were from different basic-level categories (e.g., a convertible and an airplane). The description used either a basic-level or a subordinate term. Hence, there were four critical true trials for each set (all labels referring to the convertible in this example):

1. Related pictures, basic level name – convertible and limo, “car”
2. Unrelated pictures, basic level name - convertible and boat, “car”
3. Related pictures, subordinate level name – convertible and limo, “convertible”
4. Unrelated pictures, subordinate level name - convertible and boat, “convertible”

Each participant saw each target item 4 times, once in each condition, for a total of 96 trials. An equal number of foils were created. These foils used the same pictures as for the critical targets, but in half of the cases, the wrong side was highlighted, and in the other half, a wrong name was provided (e.g., “jeans” when the pictures consisted of a coin and a phone). For both

kinds of false trials, half of the names were at the basic level and half were subordinates. Hence, when viewing a set of pictures, subjects did not know what type of term would be given (basic or subordinate level), nor did they know on the basis of the pictures whether their response should be yes or no.

The determiner manipulation (definite or indefinite) was a between-subjects variable, with half of the participants always hearing a definite determiner (*the*) and the other half always hearing an indefinite determiner (*a/an/ø*).

A male native speaker of British English recorded the sentence frame “Gregory, pick” and the object labels. The determiner was recorded with the object label in order to make it sound more natural. The sentence frame and the object labels including determiner were then concatenated to form a complete sentence of the form, “Gregory, pick a convertible.”

Picture norming. We tested the appropriateness of the names given to the pictures in a separate rating task (N = 17). Subjects had to indicate, on a scale of 0 to 10, how well they thought the name described the picture. Pictures were presented individually and paired with the correct basic level name, the correct subordinate level name, an incorrect basic level name (e.g., a picture of a car with the name “pizza”), and the incorrect subordinate name (e.g. a picture of a limo and the name “convertible”). Subjects saw all conditions in one of two different pseudo-random orders. The results indicated that the correct names were highly appropriate for the pictures, while the incorrect names were not: correct basic level name: 8.7 (SD = 0.9), correct subordinate name: 9.4 (SD = 0.5), incorrect basic level name: 0.2 (SD = 0.2), incorrect subordinate name: 3.4 (SD = 1.8). These results are in line with what one would expect: slightly higher ratings for correct subordinate names than correct basic level names, due to the more specific description containing more features that overlap with the picture, and higher ratings for

incorrect subordinate names than incorrect basic level names, also due to a greater overlap of features (a convertible shares more features with a limo than a pizza). Note that the incorrect pairings are only used as distractors in the experiments and do not form part of the analyses. The main takeaway of this rating task is that the names used in the experiment described the pictures well, and the subordinate names matched pictures more closely than basic names did, showing that they accurately described the pictures..

Procedure. In the present experiment, an imaginary third person (Gregory) was asked to pick an object out of two objects presented side by side. Participants first saw the objects, which then disappeared, and then they heard a sentence such as “Gregory, pick a/the car.” At the same time they saw highlighted one of the two sides where the objects had been, indicating Gregory’s choice, and their task was to say whether he had correctly followed the instruction.

Students were tested individually in a testing booth. At the start of the experiment, they read the following instructions: “You will always first see two pictures side by side for 3 seconds. They then disappear and you will hear someone say a sentence such as ‘Gregory, pick a train’ and at the same time a box will appear where one of the pictures was (either on the left or on the right). This is the object that ‘Gregory’ has chosen in response to the instruction. Your task is to indicate whether you think Gregory followed the instruction correctly or not. For example, there might have been a picture of a train on the left and a plane on the right, and the box indicated that Gregory had picked the train. In this case, you would press the YES button. If the box had been on the right, where the plane had been, you would press the NO button. Please answer as fast as possible without compromising accuracy!” This was followed by 5 practice trials and 3 blocks of 64 items, which were separated by subject-controlled breaks.

The items appeared in a pseudo-randomized order, keeping instances of the same item at

least 4 trials apart (average distance: 39.7 trials). Items were counterbalanced over two lists so that if a basic level term was used for an item in the first list, it would be paired with a subordinate term in the second list, and vice versa. In addition, each list could appear in one of two versions, with picture location reversed. The same lists were used for the indefinite and the definite determiner instruction. An equal number of subjects saw each list in each version.

Following a fixation point (500 ms), the pictures appeared for 3 s, followed by a blank screen for 1 s. Then one location was highlighted by a black square and simultaneously the instruction (“Gregory, pick a/the NOUN”) was played over external speakers. The target location was equally likely to appear on the left or the right, both for the critical and the filler items. Timing started at the onset of the square/sentence and lasted until the subject made a response or until 5 s had elapsed. No feedback was given. See Figure 1 for an example display.

Results

Table 1 shows the mean RTs and errors. RTs less than 250 ms (1.7%) were removed, and only RTs for correct responses were included in the analyses. For the error analyses, we used log-transformed percentages. The data were subjected to 2 (Determiner: indefinite vs. definite) x 2 (Relatedness: related vs. unrelated pictures) x 2 (Level: basic vs. subordinate level) ANOVAs. Determiner was a between-subjects, within-items factor, while the other factors were within-subjects and -items. The main effect of level is not very informative as differences in label length and frequency likely influenced the results (subordinates almost always being longer and less frequent than basic names). We report results of subject and item ANOVAs. We give the most credence to effects that are reliable in both but we also interpret effects that are significant in only one analysis.

The error analyses revealed a significant main effect of determiner, with a larger number of

errors for the indefinite determiner, $F_1(1, 52) = 4.07, p < .05$; $F_2(1,47) = 16.44, p < .001$, a main effect of relatedness, with a greater number of errors when pictures were related, $F_1(1, 52) = 11.12, p < .01$; $F_2(1, 47) = 7.77, p < .01$, and a main effect of level, with, surprisingly, more errors when the basic level description was used, $F_1(1, 52) = 12.64, p = .001$; $F_2(1, 47) = 7.21, p = .01$. However, that main effect was qualified by an interaction of level and relatedness, $F_1(1, 52) = 11.19, p < .01$; $F_2(1, 47) = 5.65, p < .05$. As shown in Table 1, errors were higher in the basic-related condition than in the other three conditions. This is consistent with the Gricean predictions discussed above: Basic names like *car* are more difficult when there are two cars. Planned comparisons showed a higher number of errors in the related than the unrelated basic level condition, $t_1(53) = 4.95, p < .001$; $t_2(47) = 2.98, p < .01$, but no such difference between the related and unrelated subordinate level conditions ($ts < 1$). Determiner did not enter into any interactions, $ps > .23$.

The RT analyses revealed longer RTs when the pictures were related, $F_1(1, 52) = 32.90, p < .001$; $F_2(1,47) = 12.81, p = .001$. The effect of category level was not significant, $F_1(1, 52) = 3.33, p < .08$; $F_2 < 1$. As in the error analysis, there was a relatedness by level interaction (marginal in the item analysis), $F_1(1, 52) = 16.73, p < .001$; $F_2(1,47) = 3.10, p = .085$. Planned comparisons showed significantly longer RTs for the related-basic level condition compared to the unrelated-basic level condition, $t_1(53) = 5.82, p < .001$; $t_2(47) = 3.80, p < .001$. This difference was smaller and not fully reliable for the subordinate level conditions, $t_1(53) = 1.95, p = .056$; $t_2(47) = 1.15, p > .25$. The determiner factor approached significance in the item analysis, with faster responses when a definite determiner was used, $F_1 < 1$; $F_2(1,47) = 4.02, p = .051$, but determiner did not interact with any other factors ($ps > .13$).

Discussion

The pattern of results for the critical predictions is clear: When there were two related pictures (e.g., a limo and a convertible) and a basic level term was used to describe one of them (*a/the car*), subjects made more errors and had increased RTs compared to when the two pictures were unrelated. This pattern was not found when a subordinate level term was used. These results suggest that Gricean considerations of using a name that distinguishes between items are at play when a basic level name is used.

The results also speak to the second issue - whether this effect would be exaggerated by the presence of a definite determiner, which more strongly requires distinction. There were no relevant effects of determiner in the main ANOVAs, but because this prediction is confined to basic-level names, we analyzed just those names to see whether the determiner had an effect on them in particular. We discovered a main effect of relatedness (as above), $F_1(1, 52) = 34.95, p < .001$; $F_2(1,47) = 13.66, p = .001$, no effect of determiner, $F_s < 1$, and an interaction over items, $F_1(1, 52) = 2.64, p = .11$; $F_2(1, 47) = 4.54, p < .05$. The interaction was driven by a stronger relatedness effect for the definite determiner (187 ms) compared to the indefinite (106 ms), though planned comparisons showed that for both determiners, the relatedness effect was significant (for indefinite: $p_s < .05$, for definite: $p_s < .001$). This suggests that while the relatedness effect was present for both definite and indefinite determiners for the basic level noun, the effect was more robust for the definite determiner. That is, “pick the car” was difficult when there were two cars, but “pick a car” was also difficult, though its effect was slightly less. This is only weak evidence for a Gricean effect, given that the interaction was not found over subjects.

We also noted that for subordinate labels, the definite determiner should have been preferred to the indefinite determiner, since there was only one object fitting the name. However, the

definite determiner was preferred as a whole, with no reliable interactions.

We will postpone attempting to explain these results until they have been replicated in Experiment 2.

Experiment 2

Experiment 1 used a between-subjects manipulation of determiner, which might be responsible for the weak effects of that variable. It is possible that the repetition of the same determiner caused participants to pay less attention to it and focus more on the noun. Alternatively, what effects we did find might have been caused by hearing the same determiner used over and over again, emphasizing its meaning. We therefore employed a similar design and the same stimuli as in Experiment 2, but with determiner varied within participants. The predictions of Experiment 1 also hold for the current experiment.

Subjects. Forty-seven undergraduate students (43 female) from the University of Birmingham participated in the experiment for course credit. Four participants with over 20% errors were omitted from the analyses.

Materials and procedure. The same materials and procedure were used as in Experiment 1. Items were counterbalanced over 4 lists so that each participant saw an equal number of definite and indefinite determiners in each condition, and an equal number of items per condition.

Subjects performed the same number of trials as in Experiment 1.

Results

As in Experiment 2, RTs less than 250 ms were deleted (2.7%). The same analyses were carried out as for Experiment 2, but with the factor Determiner now as a within-participant factor.

Results are shown in Table 2.

The error analyses showed significantly more errors for related than unrelated items ($F_1(1,$

42) = 14.02, $p = .001$; $F_2(1, 23) = 6.68$, $p < .05$), a main effect of level (only approaching significance in the item analysis: $F_1(1, 42) = 4.99$, $p < .05$; $F_2(1, 23) = 3.54$, $p = .07$), with again more errors to basic level items, and a significant interaction between level and relatedness ($F_1(1, 42) = 7.37$, $p = .01$; $F_2(1, 23) = 8.47$, $p < .01$). Mirroring the results of Experiment 1, this was caused by a higher number of errors for the basic-related condition. Planned comparisons showed that more errors were made on basic-related items than basic-unrelated items, $t_1(42) = 3.27$, $p < .01$; $t_2(23) = 4.06$, $p < .001$, while no difference was found between the subordinate level conditions ($ts < 1$). There were no effects of determiner.

The RT analyses indicated a main effect of determiner ($F_1(1, 42) = 37.14$, $p < .001$; $F_2(1, 23) = 20.11$, $p < .001$), with longer reaction times when a definite rather than an indefinite determiner was used, a main effect of relatedness ($F_1(1, 42) = 40.67$, $p < .001$; $F_2(1, 23) = 30.86$, $p < .001$), with related items taking longer than unrelated, and a significant interaction between relatedness and level ($F_1(1, 42) = 12.33$, $p = .001$; $F_2(1, 23) = 12.66$, $p < .01$). As before, this was caused by longer reaction times for basic-related items compared to basic-unrelated items (difference: 194 ms), while this difference was smaller (57 ms) between subordinate-related and subordinate-unrelated items, as shown in Figure 2. Thus, these results replicate the critical interaction found in Experiment 1.

In addition, we found a significant interaction between determiner and level ($F_1(1, 42) = 6.94$, $p < .05$; $F_2(1, 23) = 9.01$, $p < .01$), and a near-significant 3-way interaction in the participants' analysis ($F_1(1, 42) = 3.98$, $p = .053$; $F_2 < 1$). The determiner by level interaction was due to basic level items being on average 118 ms slower when preceded by a definite determiner ($t_1(42) = 5.03$, $p < .001$; $t_2(23) = 6.44$, $p < .001$) compared to an indefinite determiner, while this difference was only 48 ms for subordinate level items ($t_1(42) = 2.93$, $p < .01$; $t_2(23) =$

2.03, $p < .06$). More interestingly, and mirroring the results of Experiment 1, the marginal 3-way interaction was mainly driven by a stronger relatedness effect for the basic level items when paired with a definite determiner (233 ms), compared to a 163 ms effect when paired with an indefinite determiner (see Figure 2), though the relatedness effect was highly significant for both determiners (all $ps \leq .001$).

Discussion

The results of Experiment 2 confirmed the findings of Experiment 1, revealing a significant cost when reference is made to the basic level category name when paired with a related picture. In addition, the extra relatedness cost associated with using a definite determiner, which we found in Experiment 1, was also apparent here. However, in the present experiment we also found a cost for the subordinate term when two related pictures were presented, though it was smaller than for the basic term. While in the previous experiments there was a numerical difference for the subordinate terms, this did not reach significance.

Recall that subordinate names seem most felicitous with a definite determiner in this paradigm, as there was only one object that could receive that name in any display. In Experiment 1, there was a tendency towards faster and more accurate performance with the definite determiner, although it was not limited to subordinates. Here, when determiner was varied within subjects, that pattern disappeared, as there was a tendency for the definite determiner to be harder overall. Thus, this is not in keeping with the Gricean prediction.

General Discussion

The results were an unexpected pattern of effects consistent with some Gricean predictions along with failures of other predictions. First, the positive finding was that people found it difficult to deal with pictures from the same basic-level category, e.g., two cars, when labeling

items at the basic level, found in both errors and RTs. Presumably this was due to a potential confusion of reference (even when in fact there was no possibility of confusion as the intended target was always explicitly indicated). People would apparently prefer for such items to be distinguished even when the task does not require it. When different determiners are used, the appropriateness of the basic-level name appears to change. Saying “Pick a car” is perfectly reasonable when there are two cars present, but saying “Pick the car” is infelicitous, because it suggests that the referent is identifiable when it isn’t (by the utterance and current context). However, there was never a significant interaction of determiner and relatedness at the basic level. That is, saying “Pick a car” did not wipe out the problem of related items (although it was numerically smaller than the effect for definite determiners). This then is evidence against the claim that such presuppositions are necessarily computed and influence comprehension.

We expected *the* to be preferred for subordinates, because there was only one object fitting that name in all conditions. There is a hint of such a difference in Experiment 1 (see Table 1), but because that comparison is between-subjects, it is not the clearest comparison. In Experiment 2, definite determiners were slower overall, and there were no interactions involving determiner that would suggest different patterns between basic and subordinate levels. Thus, this does not support a Gricean account that people expect speakers to use the determiner that satisfies the most presuppositions and are confused when they do not (see Schlenker, 2012).

It is not the case that subjects are insensitive to variations in determiners. Classic studies of reading and inference have shown that readers are sensitive to determiners, though usually this is in the context of a determiner specifying a discourse referent (e.g., Haviland & Clark, 1974; Murphy, 1984). However, we found no consistent effects in a simple paradigm of picture labeling. Here the referent is already visually available, and the determiner wasn’t necessary to

answer the question. For example, if the instruction had been “Gregory, pick car,” the task could have been accomplished just as easily (except for the effect of an ungrammatical sentence). In this context, one determiner was more felicitous than the other, but this had little effect on listeners’ performance, unlike the choice of lexical item, which had a consistent effect.

Taken as a whole, it doesn’t seem that Gricean and other pragmatic principles of interpretation apply very well in this task. The effect of relatedness is consistent with the idea that one should give discriminating descriptions, but it applied even with an unambiguous referential cue (the box) and when the determiner canceled the requirement (“Pick a car”). Other predictions regarding subordinates were not found, nor were there any expected effects of determiners. Thus, we interpret the effect of relatedness as a kind of general strategy, probably with roots in the Gricean maxim, but with effects that go beyond it. That is, people generally expect object labels to be unique, because in most situations such descriptions are necessary or useful (Cruse, 1977). When a label that could fit both objects is used, people have greater difficulty with it. It is important to note that this strategy actually interferes with task performance in some conditions. For example, when the instruction is “Pick a/the car,” and there are two cars, subjects could respond positively without attending to the cue to indicate which one is picked. It makes no difference, since both are cars. When only one picture is a car, they must check that the correct one has been selected. Yet, the related condition in which both were cars was slower (and had double the number of errors), in both Experiments 1 and 2, an effect that was not removed by the indefinite determiner. Thus, the strategy of expecting distinctive names may normally be useful, but it does not seem possible to turn it off under circumstances in which it would be unhelpful.

The results are reminiscent of Engelhardt et al.’s (2006) finding that speakers produced over-

specific descriptions such as “put the apple on the towel in the box” even when there was only one apple, and they consistently avoided underdescriptions. Subjects also found overdescriptions as acceptable as concise descriptions, though eye movements did reveal a cost for them. Davies and Katsos (2013) argued that the failure to find an effect of overinformativeness was due to the complexity of the design, which would drain processing resources. However, we failed to find convincing evidence of an overinformativeness effect in a design, which was, if anything, even simpler than that of Davies and Katsos (2013, Experiment 2). In contrast, underinformativeness does cause problems for listeners, as found both in the visual world paradigm of Engelhardt et al. and Nieuwland, Ditman, and Kuperberg’s (2010) ERP study of scalar implicatures (see also Bott & Noveck, 2004).

These findings may be related to the observation of Cruse (1977) that the preference for basic-level names is not fully consistent with the maxim of quantity. In contexts in which everyone knows exactly what the referent is, if speakers are to provide enough information but no more, they should provide the least specific label necessary to pick it out. However, they do not, saying things like, “Have you taken the dog out?” even when there is only one pet. Or they may say, “My glasses are on the table” even when there is only one item of furniture in the context. The general rule for labeling seems to be that the basic level name should be used unless further specification is needed, or when the speaker is trying to bring some specific information to attention (e.g., references to “my Jaguar” or “the Doberman”).

Why is the basic level generally preferred? One might point to the fact that it is more frequent than names at other levels, but this is a circular argument in which the fact that the basic level is often used is explained by its high frequency. The question is *why* its frequency is high. Furthermore, people prefer the basic level even in cases where a more general, sufficient

description is more frequent. For example, calling something “an animal” (a superordinate term) does not seem felicitous when referring to an anteater or wolf in spite of *animal*’s much higher frequency.

The answer may be related to Schlenker’s (2012) notion of *fallibility*. Although listeners should be able to pick out referents and draw conclusions from minimal descriptions, if they have forgotten or merely stopped paying attention to part of the context, the minimal description may be difficult to compute. For example, “Have you taken the animal out?” is understandable as referring to the family’s single pet, but if one is not thinking about pets, then it might be easier to understand “dog” instead. The specificity of *table* helps one quickly comprehend the referent and find the object even if there is no other furniture in the room.

Comparison to Categorization Results

Our results may seem inconsistent with some of the production findings reviewed in the Introduction. Recall that Rosch et al. (1976) found virtually no effect of sets of stimuli, in which multiple members of categories did or did not occur. Their subjects predominantly used the basic level regardless. For the set of items in which only one member of a superordinate occurred, this finding is consistent with the overdescription as in Engelhardt et al. or our results. However, their set with two members of a basic category did not result in people using subordinate names. Lin et al. (1987) found it almost impossible to persuade speakers to use subordinate names with contextual and task requirements. However, these experiments involved labeling of individual objects, and the contextual manipulation was either across items or a prior task. Subjects were not asked to label one of multiple objects present at the same time. Indeed, when the task is to describe one of a set of objects, speakers give specific descriptions that usually allow correct identification of the referent (Clark & Wilkes-Gibbs, 1987; Engelhardt et al., 2006). Our results

suggest that listeners expect such distinctive descriptions and are slowed when they don't receive them, even when the task does not require it.

One unexpected result was that the basic level names were associated with higher errors and sometimes slower RTs, contrary to the usual basic-level advantage. Some of this effect is due to the difficulty people had with related items and basic-level names discussed above. But even in the unrelated conditions (with only one car), people did not respond noticeably faster to the basic than the subordinate name. Most demonstrations of faster basic-level categorizations use a paradigm in which the category name is provided before or at the same time as the picture (e.g., Murphy & Brownell, 1985; Rosch et al., 1976). Because subjects in the present experiments could look at the picture for some time before hearing the category label, they may have been considering potential classifications in advance, which could have eliminated any RT difference in the accessibility of the labels.

In conclusion, when people are verifying object labels, they do use some strategies that are typically found in conversation, namely, the expectation that labels will be distinctive. However, this does not seem attributable to a general use of conversational pragmatics, as the Gricean rule to be only as informative as necessary did not seem to be followed, nor did presuppositions regarding definite and indefinite determiners influence performance. Thus, we conclude that use of conversational maxims is not mandatory in this type of context and may be restricted to situations in which there are actual speakers and addressees.

Acknowledgements

We would like to thank Rebecca Bainbridge and Lotifa Ali for their help in collecting the data.

References

- Anglin, J. M. (1977). *Word, object, and conceptual development*. New York: Norton.
- Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language, 51*, 437-457.
- Brown, R. (1958). How should a thing be called? *Psychological Review, 65*, 14-21.
- Callanan, M. A. (1990). Parents' description of objects: Potential data for children's inferences about category principles. *Cognitive Development, 5*, 101-122.
- Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J.-F. LeNy & W. Kintsch (Eds.), *Language and comprehension* (pp. 287-299). Amsterdam: North-Holland.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39.
- Cruse, D. A. (1977). The pragmatics of lexical specificity. *Journal of Linguistics, 13*, 153-164.
- Davies, C., & Katsos, N. (2013). Are speakers and listeners 'only moderately Gricean'? An empirical response to Engelhardt et al. (2006). *Journal of Pragmatics, 49*, 78-106.
- Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean maxim of quantity? *Journal of Memory and Language, 54*, 554-573.
doi:10.1016/j.jml.2005.12.009
- Eskritt, M., Whalen, J., & Lee, K. (2008). Preschoolers can recognize violations of the Gricean maxims. *British Journal of Developmental Psychology, 26*, 435-443.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41-58). New York: Academic Press.

- Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, *13*, 512-521.
- Lin, E. L., Murphy, G. L., & Shoben, E. J. (1997). The effect of prior processing episodes on basic-level superiority. *Quarterly Journal of Experimental Psychology*, *50A*, 25-48.
- Murphy, G. L. (1984). Establishing and accessing referents in discourse. *Memory & Cognition*, *12*, 489-497.
- Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 70-84.
- Niewland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, *63*, 324-346.
- Okanda, M., Asada, K., Moriguchi, Y., & Itakura, S. (2015). Understanding violations of Gricean maxims in preschoolers and adults. *Frontiers in Psychology*, *6*, Article ID 901.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382-439.
- Schlenker, P. (2012). Maximize presupposition and Gricean reasoning. *Natural Language Semantics*, *20*, 391-429. doi:10.1007/s11050-012-9085-2
- Wisniewski, E. J., & Murphy, G. L. (1989). Superordinate and basic category names in discourse: A textual analysis. *Discourse Processes*, *12*, 245-261.

Table 1. Mean RTs (ms) and percentage errors (*SDs* in parentheses) in Experiment 1.

	Indefinite determiner		Definite determiner		Total	
	RT	Errors	RT	Errors	RT	Errors
Related pictures – basic level	1382 (310)	12.3	1404 (309)	11.0	1393	11.6
Unrelated pictures – basic level	1276 (372)	7.8	1217 (249)	4.2	1246	6.0
Related pictures – subordinate level	1335 (330)	6.9	1287 (277)	4.2	1311	5.5
Unrelated pictures - subordinate level	1311 (343)	6.6	1248 (270)	4.3	1279	5.4

Table 2. Mean RTs (ms) and percentage errors (*SDs* in parentheses) in Experiment 2.

	Indefinite determiner		Definite determiner		Total	
	RT	Errors	RT	Errors	RT	Errors
Related pictures – basic level	1232 (244)	8.5	1385 (353)	11.2	1308	9.9
Unrelated pictures – basic level	1069 (201)	3.7	1152 (225)	4.5	1110	4.1
Related pictures – subordinate level	1200 (185)	4.7	1230 (197)	6.0	1215	5.4
Unrelated pictures – subordinate level	1126 (181)	5.4	1191 (192)	3.7	1158	4.6

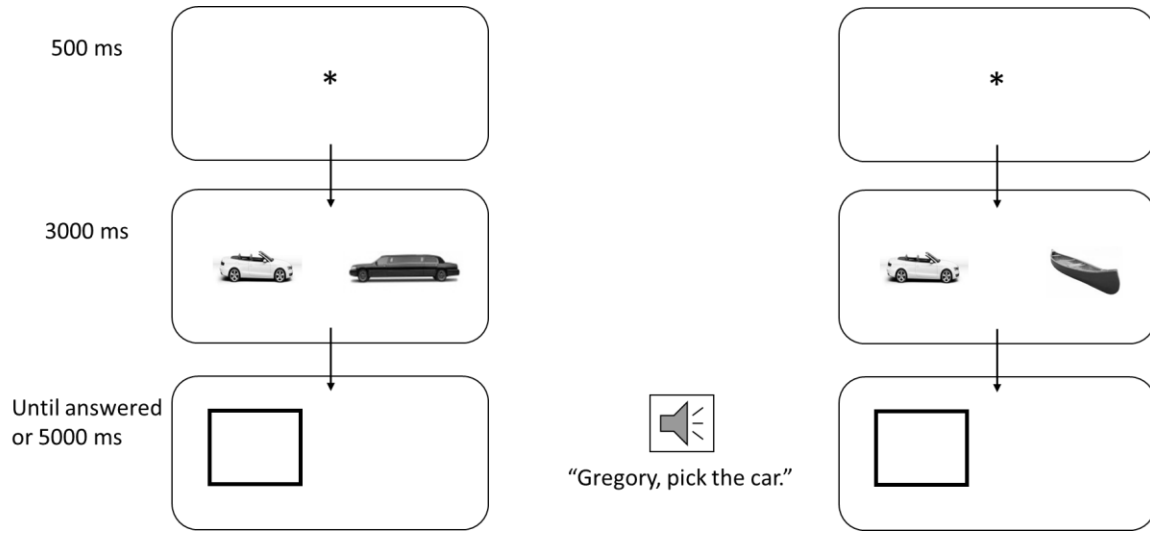


Figure 1. Example display. The example on the left depicts the related pictures – basic level name, the example on the right depicts the unrelated pictures – basic level name. For the subordinate conditions, subjects would hear “convertible” instead of “car.”

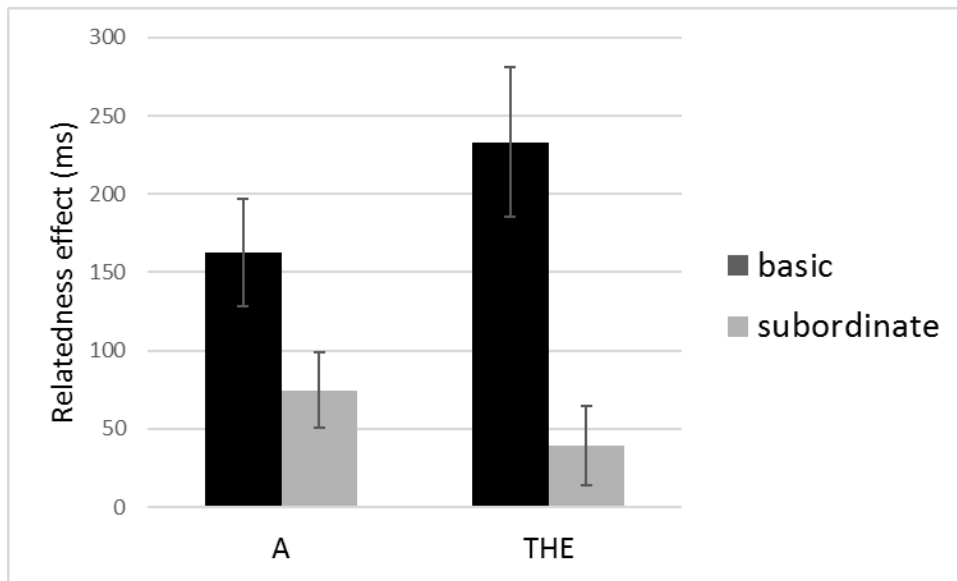


Figure 2. Difference scores (related – unrelated) with standard error bars as a function of category level and determiner, Experiment 2. The effect of related displays (e.g., two cars) was particularly noted for basic-level names. Difference scores and standard errors calculated on the participant level.