On the processing of "might"

Dimka Atanassov, Florian Schwarz and John C. Trueswell*

1 Introduction

This study examines the processing of the implicature of "might" (i.e., *not must*). The literature on implicature processing contains both studies that suggest rapid computation of scalar implicatures (e.g, Sedivy et al. 1999, Grodner et al. 2010), and studies that provide evidence for extra processing costs in generating them (e.g., Noveck and Posada 2003, Huang and Snedeker 2011). The present study extends existing work by comparing "might" to "must", and by adapting a paradigm that integrates experimental sentences into a natural discourse within a game. Our results show that the implicature does not emerge until 800ms after "might" is heard, and hence support the view that implicature processing is delayed.

In section 2 we provide a short summary of theoretical work about scalar implicatures, as well as a summary of previous psycho-linguistic work. Section 3 contains a discussion of the present study. Section 4 contains summary and concluding remarks.

2 Background

The notion of implicature was first introduced by Grice (1975), as encompassing those aspects of what is conveyed that go beyond truth-conditional meaning. Grice (1975) proposed a set of maxims, describing what it means to be an informative and cooperative speaker in conversation. Generalized conversational implicatures in particular arise via these maxims under normal circumstances. Scalar implicatures are a type of generalized conversational implicatures, which arise when a speaker uses a non-maximal value on some salient scale. Scalar implicatures arise due to the **maxim of quantity**: "be as informative as necessary, but no more than that" and in interaction with the **maxim of quality**: "be truthful". The implicature is that stronger (greater) values on the scale are either false or unknown, because if a speaker knew a stronger value to hold, he should have asserted that stronger value. For example, (1) is compatible with both meanings (1a) and (1b)

(1) Sam ate some of the chocolate cookies.
  a. Sam ate some, but not all, of the chocolate cookies.
  b. Sam ate all of the chocolate cookies.

Following Grice's approach, formal accounts have suggested that both interpretations are derived from a single literal meaning (Gadzar 1979, Horn 1984), corresponding to *some, and possibly all*. The *some, but not all* interpretation is derived by applying the **maxim of quantity** to a scale of linguistic expressions, ordered by strength in terms of entailment (Gadzar 1979, Horn, 1984). "All" is stronger than "some" on the relevant scale, as the meaning of the former unilaterally entails the latter. Therefore, in general, if a cooperative speaker was in a position to assert a strong statement, e.g., one involving "all", he would have done so (in order to be maximally informative). The fact that he used "some" implies that he was not in a position to make the stronger statement. Hence, "some" is interpreted by the hearer as *some, but not all*, via a scalar implicature. Further evidence that this implicatures is not part of the literal meaning is that it can later on be explicitly cancelled, as in (2).

(2) A: What happened to the chocolates I left in the fridge?
  B: Well, Bill ate some of them. In fact, he ate all of them.

A similar line of reasoning can be applied to a large number of expressions (Horn 1984), including "might" vs. "must". A statement of the form *"Kate must be in her office"* entails *"Kate*

* Many thanks to our confederates: Aviad Eilam and David Faber

might be in her office" (but the entailment does not hold in the reverse direction), and hence "must" is more informative than "might". Like "some", "might" implies that the speaker was not in a position to use the stronger "must". For example, in (3) it is implied that the speaker is not in a position to make the stronger claim that Kate must be in her office (if he was, via the reasoning above, he would have done so).

(3) A: Would you happen to know where Kate is?
B: She might be in her office.

As with "some", this implicature can be explicitly cancelled later. This is shown in (4) which is analogous to (2):

(4) A: Would you happen to know where Kate is?
B: She might be in her office. In fact, she must be in her office. She's always in her office on Wednesdays.

While the original theoretical debates for the most part did not make any specific claims as to the actual cognitive processes involved in implicature interpretation, there is an ongoing debate in the literature on psycho-linguistics and experimental pragmatics on the online processing of implicatures. Some initial psycholinguistic studies using eye tracking suggest that (at least certain types of) implicatures are computed rapidly (Sedivy et al. 1999), but others, using reaction time measurements, suggest that scalar implicatures are delayed (Bott and Noveck 2004, Noveck and Posada 2003). Recent studies utilizing the visual world paradigm have yielded somewhat mixed results as well. In one series of studies, Huang and Snedeker (Huang and Snedeker 2011, Huang and Snedeker 2009) offer evidence that the processing of implicatures takes additional time. Grodner et al. (2010), on the other hand, offer evidence in the opposite direction: namely that scalar inferences arise immediately, and suggest that the delays found in other studies are caused by integrating the interpretation of the inferences with contextual information.

The visual world paradigm, which is also used in the study reported here, seems well-suited for investigating the emergence of interpretations in online processing. In a typical setup, subjects face a visual display while being exposed to auditory stimuli. Measuring eye movements relative to the onset of critical expressions in the linguistic stimuli has proved to provide a fine-grained measure of the time course of the generation of interpretation of linguistic input (Tannenhaus et al 1995).

Using this paradigm, Huang and Snedeker (2011) argue that implicatures take about 600 additional milliseconds to process. The study compared the processing of "some of" to that of "two/three/all of". Participants were presented with sentences of the form: *"Point to the girl that has some/all/two/three of the ice cream sandwiches"*. For *some* and *all* trials a set of 4 items was evenly split between a boy and a girl character that were horizontally adjacent, and another set of 3 items was given to one of the remaining characters (there was a total of 4 characters). That is, the display corresponding to the above utterance contained a girl with 2 of 4 ice cream sandwiches (the target for *some* trials), a girl with 3 out of 3 ice cream cones (the distractor for *some* trials), a boy with the remaining 2 ice cream sandwiches, and a boy with no items. For the *two* and *three* trials, the first set of items was evenly split between a boy and a girl, and a second set of 4 items was unevenly split between the other boy and girl pair. The target was the character that was requested, and the distractor was the character that matched the gender but had a different item. While for the *all*, *two* and *three* conditions, the proportion of looks to the target was quick to rise (it rose above chance within the first 200-300 milliseconds), this did not happen with the *some* condition. In the *some* trials, participants developed a target preference before phonological disambiguation, but only after an 800ms delay, after which they became more likely to switch to the target for distractor initial trials than to switch to the distractor in target initial trials. Huang and Snedeker concluded that pragmatic inferences take longer to calculate than lexical semantic meaning: pragmatic inferences kicked in about 600ms after the initial semantic effect.

In a similar study Grodner et al. (2010) found no evidence for delays of scalar implicatures involved in "some", but rather found early reflexes of implicature interpretation. They argue that the delay found in other studies therefore is not caused by making the inference itself, but rather by integrating its interpretation with relevant contextual information. The design Grodner et al. (2010) used is similar to the one in Huang and Snedeker 2011, but it differs from it in several ways.[1] First, Grodner et al. (2010) changed the pronunciation by using a contracted "summa" version of "some of"; secondly, they did not use any numeral instructions, but instead added a "none of" condition (pronounced as "*nunna")*, in addition to an "all of" (pronounced as "*alla"*) condition. Grodner et al. (2010) also added a "late *summa"* condition. In this condition there were two sets of characters that had some but not all of a set of objects. Hence for the *late summa* condition the correct target could not be identified until the point of phonological disambiguation. Finally, every trial began with a prerecorded statement that described the objects in the display. For example, subjects would hear the following instructions: *"Click on the girl who has summa the balls/ alla the balloons/nunna the items."* For the *early summa/all/nunna* condition the scene consisted of three boys and three girls, such that one girl and one boy had two balls out of a total of four, one girl had all of four balloons, one boy had nothing, and another boy had two of four other objects. For the *late summa* condition again there were a boy and a girl, such that each had two of four balls, but now the four balloons were also distributed evenly between a boy and a girl, so prior to phonological disambiguation there was no way to identify the correct target. Grodner et al. (2010) found that except for the *late summa* condition, the quantifier generated increased fixations to the target 200-300ms after onset. In addition, for the *late summa* condition, about 400ms after quantifier onset fixations to the two characters that were pragmatically consistent with "some" increased.

These results are inconsistent with Huang and Snedeker 2009, 2011. Grodner et al. (2010) suggest several reasons for this. The most important one, according to the authors, is that they did not include numerals in their commands, as numerals might have reduced the felicity of "some of" as referring to upper bound targets (and thereby caused a delay in the original study). Hence they suggest that what caused the slow down in Huang and Snedeker 2009 is not the implicature processing, but rather the fact that *"some of"* sounded somewhat strange as referring to an upper-bounded subset of items, as compared to exact numbers, such as "*two/ three of"*. Since "*two/ three of"* were present in the previous study, they caused a reduction in the felicity of "*some of"*, and hence a slow down. This explanation is supported by findings reported by Degen and Tanenhaus (2012), who offer evidence for the naturalness of "some" being sensitive to the lexical alternatives available, and on the speed with which they become available. In particular, Degen and Tanenhaus (2012) show that the naturalness of "some" is lower when the relevant set is of a small size (where number terms are rapidly available), even when number terms are not explicitly used in the experiment. They also show that adding number terms to the experimental items decreases the naturalness of "some" for small sets. Finally, in an eye tracking study Degen and Tanenhaus (2012) found availability effects for small sets, and naturalness effects for large sets. In particular, when the relevant set was of a small size there was no delay of "some" compared to "all", but both "some" and "all" were processed more slowly than the numerical "two".

Up to this point, most psycho-linguistic research on the processing of implicatures has focused on quantifiers such as "some" and "all". In order to get a better understanding of the interpretation of implicatures, we should broaden the research enterprise to include other types of scalar implicature triggers, such as epistemic modals ("might" in particular). Comparing "must" to "might" has the advantage of abstracting away from specific properties of "some" vs. "all", which might have influenced the previous studies. In particular, Grodner et al. (2010) argue that the felicity of "some (of)" might be influenced by the listener hearing exact numerical expressions in the same context. This could not be a concern when comparing modals such as "might" and "must", since there are no explicit cardinalities involved in epistemic modality.

In the study described below we examine the processing of the implicature triggered by epistemic "might" and report a processing delay of 800ms. This provides further evidence for delayed scalar implicature processing, using a type of implicature trigger that has not been used in the past.

[1] In fact, the design was based on the design in an earlier study by Huang and Snedeker (Huang and Snedeker 2009).

3 A "Must" vs. "Might" Experiment

In the study presented below we examine the time course of the implicature involved in "might" (i.e, "not must").

The study incorporates the visual world paradigm into a real time conversation that takes the form of a 'guessing game' with a confederate. Incorporating the visual world paradigm into a game of guessing allowed for a more naturalistic, conversation like setting (compared to prerecorded stimuli), with two participants conversing naturally within a well-defined task-structure.

3.1 Materials and Design

The experimental design was inspired by the designs in Brown-Schmidt et al. (2008), Brown-Schmidt (2009), and Keysar et al. (2000). The participant and the confederate played a game of guessing on a computer screen. Each saw a 3x3 grid of colored shapes in which some of the shapes were occluded for one of the participants (the guesser), who had to guess what they were, aided by two simple rules. The other participant (the verifier) could see all shapes and had to respond to the guesser's guesses by marking them as "correct" or "incorrect" (depending on whether the guess matched the hidden shape in concern). The subject and confederate took turns playing the guesser and verifier: at first the subject was making guesses, while the confederate verified, and in the second part the subject and the confederate switched roles. All critical trials took place when the confederate was making the guesses. The first part (in which the subject guessed) was done to familiarize subjects with the rules of the game, as well as to make the game more believable. While for filler trials the confederate made spontaneous utterances, the target trials were partially scripted, as the confederate's screen contained written instructions on what to say. When the confederate was guessing the subject knew which parts of the scene were privileged, i.e., he was aware that these shapes could only be seen by him.

For the duration of the experiment, the guessers' guesses were recorded via a microphone, and the verifiers' eye gaze was recorded via an eye tracker. Only the second part of the experiment, during which the confederate guessed, contained the critical trials. During critical trials the guesses were phrased using either "must" or "might", depending on whether the confederate could infer his guess with certainty based on the rules.

Participants were told that there are four possible shape types: hearts, circles, triangles and squares, and four possible shape colors: red, green, blue and black. Participants were also told that the shapes will always be arranged to follow two simple rules (see also Figure 1):

Rule 1: In each row either all shapes will have the same type or all shapes will have different types.

Rule 2: In each column either all shapes will have the same color or all will have different colors.

All displays in the experiment adhered to these rules. For example, in the display shown in Figure 1 all shapes in the top row are of the same type (squares), and in the middle and bottom rows all shapes are of different types (a heart, a circle and a triangle in the middle row, and a square, a heart and a triangle in the bottom row). As for the columns, in the leftmost and rightmost column all shapes have the same color (red), and in the middle column all shapes have a different color: blue, green and black.

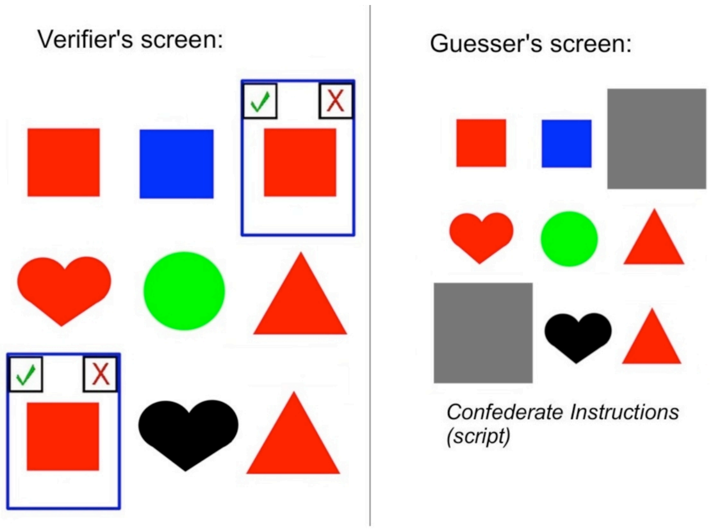

Figure1: example of a target trial. Confederate (playing 'guesser') sees right side and subject ('verifier') sees the left side. The confederate's screen also contains partial instructions on what to say for target trials.

In each trial, the subject and confederate were given five seconds to look at the screen, after which a sound was played and the person playing the guesser could start making guesses. This was done to allow the subject to familiarize himself with the scene. In addition, in order to stimulate active, rather than passive listening to the confederate's guesses, the subjects were told that they should try and respond to the guesses as quickly as possible.

Target items consisted of 12 sets of partially scripted utterances, corresponding to 12 scenes. The confederate read off his screen the scripted utterances while pretending that these were spontaneous guesses. The utterances referred to one of two hidden shapes in each scene. In the example above (Figure 1), the confederate uttered one of the following two sentences, corresponding to one of two conditions (*must* and *might* condition): [2]

1. There **must be** a red square located in the upper right.
2. There **might be** a red square located in the bottom left.

Whenever "must" was used the location referred to a shape that the confederate could guess with certainty (in Figure 1 the shape on the upper right has to be a square, because the two other shapes in the same row are squares, and it has to be red, because all other shapes in the same column are red, therefore, it can be guessed with certainty). Whenever "might" was used, the location referred to a shape that could not be guessed with certainty (in Figure 1 it can be inferred that the shape on the bottom left is red, because all other shapes in the same column are red, but the shape type cannot be inferred with certainty, as it can be either a square or a circle). Whenever "must" was used the guess made matched the hidden shape that was referred to (the confederate was "right"). For half of the cases in which "might" was used the guess made matched the hidden shape that was referred to, and for the other half the guess did not match the shape (while it was consistent with the rules). This manipulation was done so that the confederate is not surprisingly "right" all the time for cases in which guesses cannot be made with certainty.

The target sentence was preceded by an introductory sentence of the form *"Hmm/ Lets see, in the [row column] there is a [color shape]"*. This sentence always referred to a non-hidden shape, that was relevant to making the guess. That shape was always of the same type and color as the

[2] The "located" word was added to extend the ambiguity period. For some of the trials the confederate omitted the "located" word (i.e., uttered "There **must/might be** a red square in the upper right/bottom left"). Since in these trials the ambiguity period was long enough, despite this omission, the trails in concern were kept.

one that could be guessed with certainty. In the example above, the introductory sentence was *"Hmm, in the upper left there is a red square*".[3]

Each subject saw each target visual display either in the *must* or in the *might* condition (but never in both conditions), and each subject saw 6 target displays in the *must* condition and another 6 in the *might* conditions.

3.2 Predictions

There were two hypotheses to be considered. The first hypothesis (Hypothesis 1) is that pragmatic implications of "might" are directly accessed (i.e., there is no processing delay). If this is the case, then the prediction is that hearing "might" should block eye movements to the hidden shape that could be inferred with certainty (the "must" shape), and trigger eye movements to the shape that could not be inferred with certainty (the "might" shape) immediately. The second hypothesis (Hypothesis 2) is that the implicature is delayed. If this is the case, then this predicts a delay in looks to the "might" shape.

3.3 Participants

14 undergraduate students participated in this study for course credit. All participants were native English speakers and undergraduate students at the University of Pennsylvania. Two participants were dropped from the analyses reported below, due to high track loss percentage.

3.4 Apparatus

The experiment was implemented using the Experiment Builder software package by SR research. The images were presented on a 17in. Samsung screen with 1680x1050 resolution, which was divided into two halves, such that the participant viewed one half and the confederate the other half. The division was made using a partition that did not allow the subject to view the confederate's screen and vice versa. The audio recording was made using a headset with a microphone. The onset of the recording was synchronized with the display of the scenes. Right eye gaze was recorded using an Eyelink 1000 eye tracker on a desktop mount at a sampling rate of 1kHz.

3.5 Results

We initially examined the time window between 500 milliseconds before the onset of the modal and up until 2 seconds after the onset of the column indicator. The target was taken to be the shape that could be guessed with certainty (corresponding to the "must"), and the competitor the shape that could not be guessed with certainty (corresponding to the "might"). There are two conditions: *must*, corresponding to when the subject heard "must", and *might*, corresponding to when the subject heard "might". Target advantage score was computed as looks to the target minus looks to the competitor. This measure is appropriate since it shows most clearly the preference of the target over the competitor.

Figure 2 shows the target advantage score from 500ms before the onset of the modal and until 2 seconds after the onset of the column word (i.e., for the duration of the target sentence), for the two conditions: *must* and *might*. This figure shows that around 1000ms after the modal onset there is a clear divergence between the *must* and the *might* conditions. At 2000ms (2 seconds) after modal onset the divergence becomes even bigger. This graph then can serve as a sanity check, as when the location is revealed (on average at 1980ms after modal onset), the target advantage score for the competitor vs. target diverge appropriately.

[3] For the first three subjects the introductory sentence that was used was slightly different: the shape and color were specified prior to the location, the confederate uttered: *"Hmm, there is a red square in the upper left"*. This was changed for later subjects to the version above, in order to avoid subjects mistakenly responding to that sentence as the guess.

Figure2: Target advantage (must- might) starting from 500ms prior to modal onset until 2 seconds after column onset: around 1000ms after the modal onset there is a clear divergence between the *must* and the *might* conditions.

The ambiguity period was defined as starting from the modal onset ("must" or "might"), and until 200ms after the onset of the row locator. Figure 3 shows the target advantage for this period, for items for which the guess made was matching in the *might* condition (half of the items).[4]

[4] An effect of condition was only present in those items that had matching *might* guesses. When the guess was not matching for the *might* condition, the shape type that was guessed matched the one in the target position (color matched both the target and the competitor). That is, the shape that could be guessed with certainty (the 'target' shape) matched the one that the confederate uttered. That shape was also occluded, and thus also in the participants privileged ground. Thus it may be the case that when the guess was not matching the subjects simply looked at the shape that matched the one they heard (which happened to be the "must" shape) and was in their privileged ground. On average, the shape onset started 869ms after the modal onset. Since 1000ms is when the effect of the modal type begins to become evident, having a 'non matching shape type' guess could easily mask such an effect, explaining why it would be present only when the guess is matching.

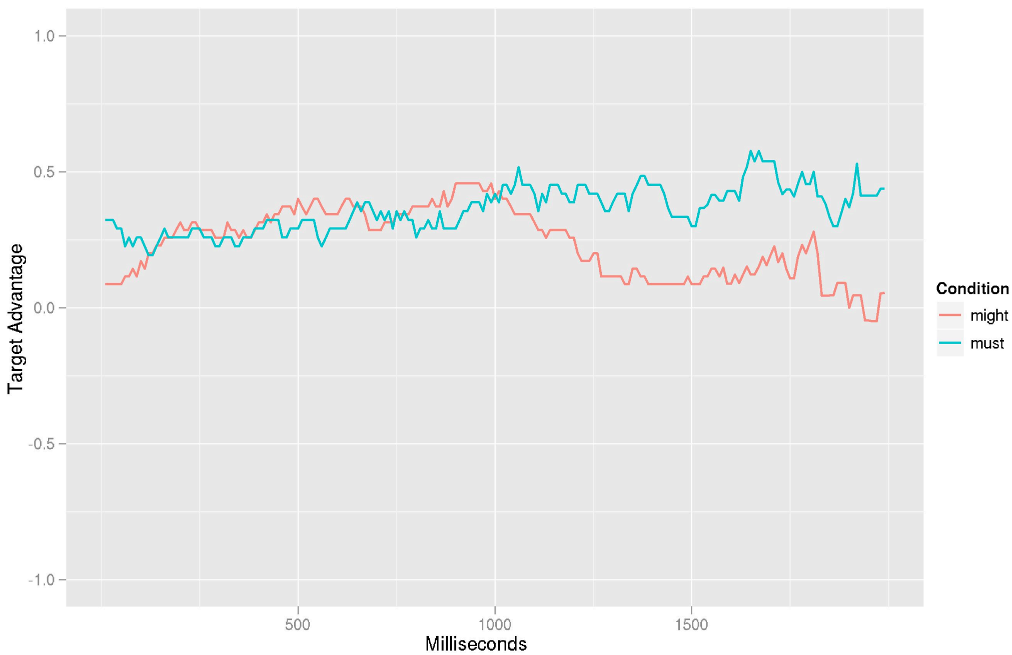

Figure 3: Target advantage (*must- might*) starting at the modal onset and until 200 milliseconds after the onset of ‘row’ for items with matching guess.

Starting from 1000ms after modal onset the *must* and *might* conditions diverge, as the target advantage score for the *might* condition lowers. Before that, the target advantage score for both conditions is (equally) high, meaning that participants looked more at the shape that could be guessed with certainty for both conditions. Thus, subjects initially looked at the target (the shape that could be guessed with certainty) more than they looked at the competitor (the shape that could not be guessed with certainty). Upon hearing “must” subjects kept looking at the target, and 1000ms after hearing “might” looks to the competitor increased relative to looks to the target. If we take the display in Figure 1 as an example, this would mean that subjects initially looked more at the target (the shape in the upper right, which could be guessed with certainty). When the confederate uttered *“There must be a red square located in the upper right”*, upon hearing “must”, participants kept looking more at the target; when the confederate uttered *“There might be a red square located in the bottom left”*, 1000ms after the onset of “might” participants started looking more at the competitor (the shape in the bottom left, which could not be guessed with certainty).

In the *might* condition subjects began directing their looks towards the appropriate shape only about 1000ms after the modal onset, and prior to that, they were looking at the shape that could be guessed with certainty. Taking into account that about 200ms are necessary to plan and perform a saccade, this can be interpreted as an 800ms delay in looks to the appropriate shape (given the implicatures) for the *might* condition.

For items with matching *might* guesses the total ambiguity period was divided into two windows: one from the modal onset and until 1 second after the modal onset (window A), and the other from 1 second after the modal onset and until 200 milliseconds after the onset of the row locator (window B). ANOVAs were done on the target advantage scores using these windows on subject and item means. There was a significant interaction between time window and condition, both by subjects ($F(1, 11)= 6.98$, $p < 0.05$) and by items ($F(1,5)= 33.47$, $p < 0.01$).

3.6 Discussion

About 1 second after the modal onset, target advantage scores for the *might* and *must* conditions diverged. Prior to that, the score for the *might* items patterned with that for the *must*. This result can be interpreted as evidence that implicatures processing is slow: prior to processing the implicature, *might* patterned with *must* and diverged from it when the implicature was processed. Taking into account 200ms for planning and performing a saccade this results in 800ms for processing the implicature in “might”.

When interpreting the results, however, one has to take into account that there is an initial target bias. Throughout the entire ambiguity period the target advantage for both conditions is above zero. This is true for both when the guess is matching and when it is not (for the *might* condition). There seems to be an initial preference to look at the shape that could be guessed with certainty even for the *might* condition, even when there was no independent reason to prefer this shape. This preference was larger when “must” was uttered, but it was nevertheless present for the *might* condition as well. Furthermore, during the first second after modal onset, there was no significant difference between *might* and *must*, and for both there was a preference to look at the shape that could be guessed with certainty. A possible explanation for the latter fact is that since the shape that could be guessed with certainty was easier to guess, the subject might have anticipated that this shape would be guessed first, and hence this shape attracted more initial looks. The items for which the *might* guess was not matching add further to the target bias. The target advantage score for the *might* conditions in which the guess was matching drops around 1000ms after the modal onset. For items in which the guess was not matching the drop in target advantage for the *might* condition is slower, and patterns with that for the *must* condition until about 2 seconds after the modal onset. As mentioned above, when the guess for the *might* condition was not matching, the shape that was guessed matched the “must” shape (i.e., the shape that could be guessed with certainty), and hence this resulted in increased looks to the target for the *might* condition in items in which the guess did not match. This initial target bias makes it more difficult to show that there was no delay in the processing of “must”, since if there is an independent initial preference to look at the “must” shape, it would be difficult to tell when looks to this shape are prompted by hearing “must” and when they just follow from an independent preference. Independent evidence that “must” is processed immediately is therefore needed.

An additional issue with the above results is the difficulty of the task. In order to attribute the effect found to the implicature involved in “might”, other factors (such as difficulty of the task) need to be ruled out. The task subjects needed to perform was fairly complicated: they needed to be able to do the kind of reasoning (using the game rules) that would allow them to realize that one shape could be guessed with certainty while the other shape could not. To aid them with this task, before the target trials took place there was a “training” period, in which subjects were given similar displays and had to use the game rules to make guesses (there were 16 such trials for each subject), and while we believe that this indeed helped subjects prepare for the task in target trials, the extent to which people seemed to absorb the rules varied across subjects (some people did better than others when playing the “guesser” themselves). Taking this into account, and since there was an effect of the condition (whether the confederate said “must” or “might”) on looks, it is reasonable to assume that people learned the game rules to the extent to which the experimental task yielded meaningful results.

4 General Discussion

Our results suggest an 800ms delay in the processing of the implicature involved in “might”. This is consistent with the results in Huang and Snedeker (2011) and Huang and Snedeker (2009). Since in this study we used modals instead of quantifiers, this allowed us to abstract away from particular problems “some” and “all” might have (such as varying felicitiousness in the presence of numeric expressions). In addition, our task was embedded in a discourse with a confederate, hence allowing for a more natural environment (compared to prerecorded stimuli). It still remains unclear how to reconcile the present results, and the results in Huang and Snedeker (2011) and Huang and Snedeker (2009), with the results in Grodner et al. (2010). There is no doubt that more data needs to be collected and more experiments conducted in order to settle with certainty the question of whether or not implicature processing is slower, and whether or not this is a matter of context or a general property of implicatures. Our results however provide further evidence that implicature processing is slow, and semantic analysis precedes pragmatic inferences.

References

Bott, Lewis and Ira A Noveck. 2004. Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language* 51: 437–457.

Brown-Schmidt, Sarah. 2009. The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review* 16:893–900.

Brown-Schmidt, Sarah, Christine Gunlogson and Michael K. Tanenhaus. 2008. Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition* 107:1122–1134.

Degen, Judith and Michael K. Tanenhous. 2012. Naturalness of lexical alternatives predicts time course of scalar implicatures. Poster presented at CUNY 25, CUNY Graduate Center, New York.

Gadzar, Gerald. 1979. *Pragmatics: Implicature, Presupposition, and Logical Form*. New York: Academic Press.

Grice, Paul. 1975. Logic and Conversation. *Syntax and Semantics* 3: 41–58.

Grodner, Daniel J., Natalie M. Klein, Kathleen M. Carbary and Michael K. Tanenhaus. 2010. “Some,” and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition* 116:42–55.

Horn, Laurence. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In *Meaning, Form and Use in Context*: *Linguistic applications*, ed. D. Schiffrin 11–42. Washington D.C.: Georgetown University Press.

Huang, Yi Ting and Jesse Snedeker. 2009. Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology* 58:376–415.

Huang, Yi Ting and Jesse Snedeker. 2011. Logic and Conversation revisited: Evidence for a division between semantic and pragmatic content in real time language comprehension. *Language and Cognitive Process*. 26:1161–1172.

Keysar, Boaz, Dale J. Barr, Jennifer A. Balin, and Jason S. Brauner. 2000. Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science* 11: 32–38.

Noveck, Ira A. and Andres Posada. 2003. Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language* 85:203–210.

Sedivi, Julie C., Michael K. Tanenhaus, Craig G. Cahmbers, and Gregory N Carlson. 1999. Achieving incremental semantic interpretations through contextual representation. *Cognition* 71:109–147.

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard and Julie E. Sedivy.. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268:1632 – 1634.

Dimka Atanassov, Florian Schwarz
Department of Linguistics

John C. Trueswell
Department of Psychology

University of Pennsylvania
Philadelphia, PA 19104-6305
dimka@ling.upenn.edu
florians@ling.upenn.edu
trueswel@psych.upenn.edu