This article was downloaded by: [K.U.Leuven - Tijdschriften]

On: 12 March 2012, At: 10:53

Publisher: Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

The Quarterly Journal of Experimental Psychology

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/pqje20>

Some effort for some: Further evidence that scalar implicatures are effortful

Kristien Dieussaert ^a, Suzanne Verkerk ^a, Ellen Gillard ^a & Walter Schaeken ^a

^a Psychology Department, University of Leuven, Leuven, Belgium

Available online: 23 May 2011

To cite this article: Kristien Dieussaert, Suzanne Verkerk, Ellen Gillard & Walter Schaeken (2011): Some effort for some: Further evidence that scalar implicatures are effortful, *The Quarterly Journal of Experimental Psychology*, 64:12, 2352-2367

To link to this article: <http://dx.doi.org/10.1080/17470218.2011.588799>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Some effort for some: Further evidence that scalar implicatures are effortful

Kristien Dieussaert, Suzanne Verkerk, Ellen Gillard, and Walter Schaeken

Psychology Department, University of Leuven, Leuven, Belgium

Under the assumption of the principle of cooperation (Grice, 1989), a statement such as “some eels are fish” is thought to be false since it contains less information than is considered sufficient. However, the statement is logically sound since the meaning of “some” is compatible with “all”. Currently, the primary interpretation of such underinformative statements remains subject to debate. According to Levinson (2000), the pragmatic “some but not all” interpretation is the default interpretation, while others (e.g., Sperber & Wilson, 1995) argue that this pragmatic interpretation only comes to the fore when relevant within the context and is thus considered secondary to the logical “some and perhaps all” interpretation. In this study, three factors that may influence the answer pattern are studied: task load, working memory capacity, and repetition of the statements. In Experiment 1, we used a secondary task paradigm to manipulate the cognitive load under which a number of underinformative statements had to be judged. We observed that for participants with a rather limited working memory span it is harder to reach a pragmatic interpretation under cognitive load. In Experiment 2, we manipulated the repetition of the statements. We observed that with a higher number of filler statements, participants produced fewer consistent answer patterns. This study provides further evidence against the automaticity of the pragmatic interpretation: It shows that the pragmatic interpretation requires more cognitive effort than the logical interpretation and that increasing the number of filler statements inhibits the development of a response strategy.

Keywords: Scalar implicatures; Working memory capacity; Logic; Pragmatic.

In daily communication, the messages that we want to convey are to a large extent not explicitly expressed. Instead, people in conversation make use of facial expressions, gesticulation, and the (assumed) intentions of the speaker to make their interactions successful. For example, if someone

were to tell a colleague that some of his children are blond, it would not be surprising that the colleague interprets this information as if not all of them are. Nevertheless, the semantic meaning of “some” is “some and perhaps all”. How then can we explain the intuitive ease with which a “some

Correspondence should be addressed to Kristien Dieussaert, Psychology Department, University of Leuven, Tienestraat 102, 3000 Leuven, Belgium. E-mail: kristien.dieussaert@psy.kuleuven.be

We are grateful to the participants of the St. Hubertuscollege and the University of Leuven. This research was conducted with the financial support of the Odysseus programme (granted to Igor Douven) and the Flanders Research Fund (FWO; project funding granted to Walter Schaeken). We also wish to thank two anonymous reviewers and Wim De Neys for commenting on a previous version of this manuscript.

but not all” interpretation is given to this sentence? Grice (1989) provided a theoretical framework to discuss this question and introduced the notion of “implicature” in linguistics. The meaning of an expression is “what is said” plus “what is implicated”. The latter refers to what the speaker means, but not explicitly says. To establish effective communication, four maxims (quantity, quality, relation, manner) should be obeyed, based on a general principle of cooperation, which Grice (1989, p. 26) formulates as follows: “Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.”

Grice (1989) distinguished between “conventional” and “conversational” implicatures. Conventional implicatures, on the one hand, are context independent and do not rely on the maxims of conversation. For example, in the statement “She is poor but honest”, the word “but” refers to a contradiction in the mind of the speaker between being poor and being honest, although the speaker does not explicitly say that he thinks there is a contradiction between these two notions. Conversational implicatures, on the other hand, are dependent on the context and do rely on the maxims of conversation. They are related to the content and not to the form of the expression (as in the case of “but”, for example) and can be withdrawn in the course of the conversation by the context or the speaker. Within the group of conversational implicatures, Grice makes a further distinction between “generalized” and “particular” implicatures. The former are conversational implicatures that are so widespread within a language community that they have become the “default meaning” and are thus less dependent on the context than the particular implicatures. Compare, for example, the following utterances:

1. He goes to the party with a woman.
2. Everybody goes to the party.

Since the speaker in (1) uses the word “a”, we immediately infer that the woman is not his wife. This is clear without any additional contextual

information. However, this is still not a conventional implicature, since it can easily be withdrawn if the speaker were to state later on that he has no knowledge of the marital status of the man. To understand the meaning of Statement 2, more contextual information is required. One possible situation may be that a young speaker is arguing with his father about going to that party and is trying to say “why can I not go?”. Another possible context could be that the speaker wants to convince the hearer to also come to the party since it will be fun with so many people around.

In this article, we deal with a special type of generalized conversational implicatures: the scalar implicature. This concept was developed by Horn (1972). Horn-scales consist of a set of alternative expressions of the same grammatical category, but with a difference in semantic informativeness. The general assumption underlying these scales is that the use of a semantically weaker term implies that the stronger one does not hold. This assumption is in accordance with the maxim of quantity (“be as informative as required”, “be not more informative as required”). These are a few examples of such scales, ordered from strong to weak:

<all, most, some>
 <and, or>
 <certain that p, probable that p, possible that p>
 <hot, warm>
 <. . . , six, five, four, three, two, one>

Accordingly, when a speaker states that most students get their degree in June, the hearer infers that not all students get their degree in June. Otherwise, the speaker has not been as informative as required.

While the four maxims primarily served as a norm for effective communication in Grice’s (1989) work, later research focused more on how people actually deal with these implicatures in daily conversation. With this change in perspective, the maxims entered the domain of pragmatics and cognitive sciences. Currently, we can roughly distinguish two opposing views on implicatures among the researchers working in the

tradition of Grice (see Noveck & Reboul, 2008): the neo-Gricean view and relevance theory. The neo-Gricean view (e.g., Chierchia, 2004; Horn, 1984, 2004; Levinson, 1983, 2000) is grounded in the field of linguistics. Although the researchers listed above have rather diverging opinions on how implicatures work, a commonality in their views is that they all consider the pragmatic interpretation of “some” to be the primary interpretation. Within the neo-Gricean view, the most important contribution regarding *scalar implicatures* was made by Horn (1984, 1989/2001). He introduced the Q-principle (make sure your contribution is efficient; say as much as possible), which established that the use of a weaker term excludes the use of a stronger term. While Chierchia deals with scalar implicatures from a grammatical point of view, Levinson (2000) adopts a more cognitive perspective. He reinterprets the Gricean maxims as heuristics and claims that we need to resort to heuristics because we deal with incoming information very quickly, while the production of speech occurs much more slowly. To overcome this discrepancy, we rely on metalinguistic aspects to put more information in an expression than what is actually said. For example, if a cooperative speaker uses a weaker term of a semantic scale, such as in “some of her children are blond”, the listener assumes that the speaker has no reason to believe that “all” holds, or perhaps even knows for a fact that “not all” holds. The generalized conversational implicature “some but not all” arises automatically, independently of the context. This pragmatic interpretation is cancellable, albeit with some effort. Note that to our knowledge, Levinson is the only author within this view that holds such a strong claim on the automaticity of the implicature.

Relevance theory (Carston, 1998, 2004; Sperber & Wilson, 1995; Wilson & Sperber, 2004), on the other hand, approaches implicatures in a completely different way. This theory is grounded in the field of cognitive sciences. Sperber and Wilson agree with Grice (1989) that the interpretation of

an utterance is also determined by the expectations the hearer has regarding the goal of the speaker. However, they do not accept Levinson’s (2000) claims that people reach their interpretation by maxims and heuristics. Sperber and Wilson consider “relevance” to be the central notion of the human cognitive system. Our perceptual mechanisms, memory system, and inference techniques are developed to identify the most relevant elements in a particular context. Relevance is determined by two parameters: “cognitive effect” (positive) and “cognitive cost” (negative). According to their line of reasoning (see also Noveck, 2001), a (scalar) implicature will only be drawn when it is considered relevant within that particular context. For the sake of clarity, we want to stress that relevance theory does not claim that the logical interpretation is a default interpretation with no cost involved. It only claims that in utterance interpretation, the lexical meaning generally suffices and is less effortful to construe than the pragmatically enriched reading.

Let us return to Example 1 of the implicature above. According to Levinson (2000) the implicature that “a woman” is not the subject’s wife would be drawn automatically. However, if the hearer is aware that the speaker has no knowledge of the marital status of the man and only wants to offer the information that “a woman” was present, it seems quite clear that the hearer might not draw the implicature at all, as it is not considered relevant. That is the key difference between the neo-Gricean theories and relevance theory. Relevance theory states that an implicature is only drawn if relevant within the context, while the neo-Gricean view entails that the implicature is drawn by default, but can be withdrawn in the light of more information. Hence, the neo-Gricean view is also referred to as “defaultism”, while relevance theory is also known as “contextualism”.¹

Although the defaultist view is intuitively the most appealing to many people, it is the contextualist view that is most supported by empirical data. Many experiments focus on the scale <all, some>

¹ Note that the term “generalized implicature” is not applicable in the contextual view since all implicatures are context driven.

since it has the practical advantage of having a standard logical norm of correctness.² An underinformative statement, such as “some students get their degree in June”, can either be logically interpreted as “some and possibly all students get their degree in June”, or it can be pragmatically interpreted as “some but not all students get their degree in June”. The defaultist view is that the pragmatic interpretation is automatic (heuristic, fast) and that it requires cognitive effort (analytic, slow) to be withdrawn. The contextualist view is the opposite: Construing the logical interpretation is less effortful than construing the pragmatic interpretation.

Experiments with children show that they favour the logical interpretation over the pragmatic interpretation (Noveck, 2001), unless the relevance of the pragmatic interpretation is made clear (Feeney, Scafton, Duckworth, & Handley, 2004; Guasti et al., 2005; Papafragou & Musolino, 2003; Pouscoulous, Noveck, Politzer, & Bastide, 2007). These results are mostly taken as indirect evidence for the contextualist view: Since children have a more limited cognitive capacity than adults, they will prefer the interpretation with the lowest cognitive cost—that is, the logical interpretation. Another form of indirect evidence for the contextualist view is provided by latency data. Generally, it takes more time to give a pragmatic response to an underinformative statement than a logical response (Bezuidenhout & Morris, 2004; Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; Noveck & Posada, 2003). The argument here is that giving a pragmatic response requires more cognitive effort and therefore takes more time than a logical response.

Bott and Noveck (2004, Experiment 4) also provide more direct evidence for the contextualist view. By manipulating participants' response time, they observed a decrease of pragmatic responses under time pressure. De Neys and Schaeken

(2007) manipulated the load on the cognitive resources directly through the introduction of a dual task. If working memory is loaded, less capacity is available for tasks that require controlled attention. Therefore, the participant will rely more on automatic, heuristic processes rather than going through an analytic thinking process. (For more on the relation between working memory capacity and higher cognitive processes, see Baddeley, 1996; Daneman & Carpenter, 1980; Engle 2002.) De Neys and Schaeken (2007) show that experimentally limiting the working memory space available results in fewer pragmatic responses, as predicted by contextualism. However, as Feeney et al. (2004) have pointed out, these results may be mediated by the interindividual differences in working memory capacity. It is possible that only participants with relatively low working memory capacity lack the necessary resources to perform analytic³ thinking. To gain insight into this matter, a dual-task paradigm (as in De Neys & Schaeken, 2007) combined with an assessment of working memory capacity (as in Feeney et al., 2004) seems promising. Therefore, we have performed an experiment (Experiment 1) to test the combined effects of working memory capacity and cognitive load on the type of underinformative statements given. We presented participants with more than one underinformative statement. This is important, because Oberauer, Geiger, Fisher, and Weidenfeld (2007) point out that only when participants produce the same response pattern repeatedly within the same task-type can the results be attributed to their interpretation of the words rather than to random fluctuations. Neither De Neys and Schaeken nor Bott and Noveck (2004, Experiment 4) have within-participant consistency data available. By adding more underinformative statements, we had the ability to perform a consistency analysis on top of the standard response analysis. We observe highly consistent answer

² Notice that a logical norm of interpretation also exists for the frequently studied <and, or> scale, where “or” is interpreted inclusively. For other Horn-scales, such as <hot, warm>, there is no such logical norm.

³ Note that analytic thinking is not to be interpreted as “giving a logical response”. Analytic thinking refers to a process of controlled attention that require some cognitive effort. Whether that effort will result in a logical or pragmatic answer is subject to this (and previous) studies.

patterns, independently of the cognitive load put to the task. We also observe that making the task harder lowers the number of pragmatic answers. This effect is even more pronounced for participants with a low working memory capacity.

On the basis of the results of Experiment 1, we draw attention to differences in specifics of the tasks that are used in previous studies, especially in relation to the obtained answer patterns. One of the influential factors, the proportion of filler statements, is subject to further study in Experiment 2. Here also we performed a response and a consistency analysis. More interruption of the underinformative statements gives rise to a less consistent answer pattern and to fewer pragmatic answers.

EXPERIMENT 1

The aim of this experiment is to test the automaticity of scalar implicatures with a task based on previous studies such as Rips (1975) and Smith (1980). These tasks are referred to as statement evaluation tasks (Guasti et al., 2005) or sentence verification tasks (Bott & Noveck, 2004). Participants are presented with statements such as “some oaks are trees”. It is of course generally known that all oaks are trees: An “oak” is an exemplar of the category “tree”. The statement is thus underinformative, as the weaker term of the scale <all, some> is used instead of the stronger one. If a participant gives a pragmatic interpretation to the statement (i.e., “some but not all”), he will judge the underinformative statement as false. If a participant prefers the logical interpretation (i.e., “some and possibly all”), the statement will be considered true.

To distinguish between the defaultist and the contextualist view, we introduced a second task to increase the cognitive load. Placing a load on working memory is known to facilitate automatic responses and inhibit analytic responses (De Neys, 2006a, 2006b). According to defaultism, a high cognitive load will lead to more pragmatic (false) responses than a low cognitive load. Contextualism, on the other hand, predicts exactly the opposite. However, this result may be

mediated by the working memory capacity of the individual participants. Participants with high working memory capacity may still be able to suppress an automatic response, even under high cognitive load. Therefore, we also measured working memory capacity. We expect the effect of the cognitive load to be stronger for participants with relatively limited working memory capacity than for participants with higher working memory capacity.

As is standard in the statement evaluation task, we introduced a set of filler statements that were either clearly false (e.g., some spiders are trees; all trees are oaks) or clearly true (e.g., some trees are oaks; all oaks are trees). Five types of filler items were used, as shown below in Table 1. We presented a set of 10 filler statements (5 true, 5 false) to half of the participants, in order to have an equal number of underinformative and filler statements. To the other half of the participants, we presented a set of 20 filler statements in order to harmonize the number of the underinformative statements and the number of the clearly true and clearly false statements. Two types of analysis were performed on the data. First, we report the mean scores on the underinformative statements in a standard general response analysis. Second, we check the consistency of the answer pattern within each participant in a consistency analysis.

Method

Participants

A group of 106 final-year students from a secondary school in Belgium (St. Hubertuscollege, Neerpelt) took part in the experiment. The

Table 1. *Examples of underinformative and filler statements used in the statement evaluation task*

<i>Underinformative statements</i>	<i>Filler statements</i>
Some mackerel are fish	Some fish are tuna
Some oaks are trees	Some spiders are trees
Some sparrows are birds	All crows are birds
Some tulips are flowers	All flowers are roses
Some bees are insects	All flies are fish

students were between 17 and 19 years old. Female as well as male students participated.

Materials

Cognitive load task. We used a spatial memory task (Bethell-Fox & Shepard, 1988; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001) for which the stimuli and procedure were taken from De Neys (2006b). Three or four dots were shown in a 3×3 matrix for 850 ms. Participants had to remember how the dots were distributed in the matrix and reconstruct their configuration on the numerical keyboard. In the low cognitive load task, three dots were presented in one horizontal or vertical line (see Figure 1a). In the high cognitive load task, four dots were distributed over two or three rows and/or columns of the matrix (see Figure 1b).

Working memory capacity measure. To measure working memory capacity, we used a version of the Operation Span Task (OSPAN; La Pointe & Engle, 1990) that was translated into Dutch, computerized, and adapted for group testing

(GOSPAN; De Neys, d' Ydewalle, Schaeken, & Vos, 2003). This standard test mainly measures the central working memory component. It consists of remembering 15 sets of words, which are presented in serial order. Before a word appears on the screen for 800 ms, a simple maths problem has to be assessed as true or false—for example, does $(10 \div 3) + 3$ equal 13? (right answer: false). Latencies and accuracy data are automatically registered. After being confronted with a series of words varying from two to six items, participants are requested to write down the words they remember in the correct order. The number of words recalled in the correct order gives an indication of a participant's working memory capacity. The maximum score equals 60. Since it is very important that the test is completed with great concentration to obtain a valid result, the data analysis does not take into account participants who answered more than 15% of the maths problems incorrectly, as well as participants whose latencies deviated more than 2.5 standard deviation of the mean reaction time. Based on their GOSPAN scores, participants were placed in the low, intermediate, or high working memory capacity group.

Figure 1. Examples of the cognitive load task: (a) low cognitive load; (b) high cognitive load.

Statement evaluation task (SET). To measure the sensitivity of scalar implicatures, we used a statement evaluation task for the Horn scale $\langle \text{all}, \text{some} \rangle$ with exemplar–category statements (Bott & Noveck, 2004). The content of the statements always consisted of categories and exemplars (e.g., some $\langle \text{exemplar} \rangle$ are $\langle \text{category} \rangle$). Participants were requested to judge the statements as either true (logical answer) or false (pragmatic answer). We provided 10 underinformative statements (i.e., statements that are sensitive to scalar implicatures; e.g., some roses are flowers) and either 10 or 20 filler statements, which were either clearly false (e.g., some ducks are insects) or clearly true (e.g., some birds are sparrows). As shown in Table 1, statements with “all” and with “some” were included in the set (translated from Dutch), resulting in five different types of filler statements. The statements were presented in random order.

Again, we excluded participants who answered more than 15% of the filler statements incorrectly

from the data analysis. Pragmatically correct responses to the underinformative statements were coded as “1”. For each participant, we computed the mean score: A mean score close to 1 reflects a high number of pragmatic answers.

Design

The cognitive load was manipulated within subjects: Each participant performed the low as well as the high cognitive load task. Half of the participants first did the low cognitive load task and then moved on to the high cognitive load task. The other half completed the tasks in the reversed order. Accordingly, the order is a within-subjects variable. The working memory capacity of the participants was measured, and the participants were classified as having low (lowest tertile), intermediate (middle tertile), or high (highest tertile) capacity. In the analysis, this variable is considered a between-subjects variable. The number of filler statements was also manipulated between subjects. Half of the participants received 20 filler statements, while the other half only received 10 filler statements. Finally, the type of response—logical (true)—pragmatic (false)—to the underinformative statements is the dependent variable.

Procedure

The experiment was computer driven. Participants were tested in their class group and were randomly assigned to one of the various experiment conditions described in the previous section. The experiment started with the cognitive load task and statement evaluation task. The participants received the instructions of the statement evaluation task and the cognitive load task, followed by a demonstration of the cognitive load task. To make sure that both tasks were clearly understood, some practice items were included. First, two examples of the cognitive load task were presented on the screen for one second, which then had to be reconstructed on the numerical keyboard. These examples were followed by two complete practice items: A cognitive load item was shown on the screen for 850ms, after which the participants were confronted with a truth verification statement, and finally the cognitive load item had to be

reconstructed. After each practice item, the participants received feedback on the correctness of the reconstruction of the cognitive load item. After completing the statement evaluation task with high and low cognitive load, the participants started with the working memory span task. This task was also preceded by instructions and a demonstration.

Results

Exclusion criteria and basic analysis

The mean score on the low cognitive load task was 2.95 (out of 3; $SD = 0.22$), while the mean score on the high cognitive load task was 3.89 (out of 4; $SD = 0.53$). It seems safe to say that all participants performed well on these tasks. Eleven participants were excluded based on the exclusion criteria of working memory capacity mentioned above in the Materials section. The working memory capacity of the remaining 95 participants was calculated. The mean GOSPAN score was 37.41 (out of 60; $SD = 9.47$). The high working memory capacity group (highest tertile) consisted of 31 participants with a score above 41 ($M = 48.35$; $SD = 4.85$), while the low working memory capacity group (lowest tertile) consisted of 32 participants with a score below 32 ($M = 27.59$; $SD = 4.92$). The middle group consisted of 32 participants with a score between 32 and 41 ($M = 32.62$; $SD = 2.20$). An analysis of variance (ANOVA) of the log-transformed reaction times of the three groups did not reveal a difference between the three groups ($M_{\text{highest}} = 1.34$, $SD_{\text{highest}} = 0.22$; $M_{\text{middle}} = 1.28$, $SD_{\text{middle}} = 0.23$; $M_{\text{lowest}} = 1.34$, $SD_{\text{lowest}} = 0.19$). An additional participant of the lowest tertile and 2 participants of the middle tertile were excluded based on the exclusion criterion of the statement evaluation task. In the high as well as in the low working memory capacity group, 16 participants were presented with 20 filler statements, and 15 participants were presented with 10 filler statements. In the middle group, 11 participants were presented with 20 filler statements and 19 participants with 10 filler statements.

Table 2. Overview of the percentage of pragmatic answers to the underinformative statements in Experiment 1 by cognitive load and working memory capacity

Cognitive load	Working memory capacity	Pragmatic answers (%)	N
High	Low	68	31
	Middle	59	30
	High	72	31
Low	Low	78	31
	Middle	59	30
	High	73	31

Response analysis

Results are shown in Table 2. An ANOVA shows the mean effect of cognitive load on the answers in the statement evaluation task. In the high cognitive load task, the participants gave more logical answers than in the low cognitive load task (66% vs. 70%; $n = 92$), $F(1, 80) = 5.43$, $MSE = 0.01$, $p < .05$. The order in which the cognitive load task was presented has no overall effect ($n_1 = 43$; $n_2 = 49$; 66% vs. 70%). Although no general effect of the working memory capacity could be found ($n_{\text{lowest}} = 31$; $n_{\text{middle}} = 30$; $n_{\text{highest}} = 31$; 73%, 59%, and 72%), there was an interaction between working memory capacity and cognitive load, $F(2, 80) = 4.14$, $MSE = 0.01$, $p < .05$. Participants with low working memory capacity gave more logical answers under high cognitive load than under low cognitive load (68% vs. 78%), $n = 31$; $F(1, 54) = 11.19$, $MSE = 0.01$, $p < .005$. For participants with intermediate or high working memory capacity, there was no difference in the number of logical answers between the two cognitive load tasks (respectively, 59% vs. 59%, and 72% vs. 73%). No general effect of the number of filler statements could be found ($n_{20\text{fillers}} = 43$; $n_{10\text{fillers}} = 49$; 64% vs. 72%), nor was there any correlation with cognitive load or working memory capacity.

A possible criticism of the results obtained in this study and other similar experiments is that the logical answer is always “true”, while the pragmatic answer is always “false”. Unfortunately, the design does not allow us to balance out the answer types. This could lead to other possible

Table 3. Overview of the percentage of “true” answers to the filler statements in Experiment 1 by cognitive load and working memory capacity

Cognitive load	Working memory capacity	“True” for filler items (%)	N
High	Low	49	31
	Middle	50	30
	High	48	31
Low	Low	48	31
	Middle	51	30
	High	47	31

interpretations of the results. For example, answering “true” is easier than answering “false”, and therefore the number of “true” answers increases under cognitive load. In order to exclude such interpretations, we checked for shifts in the answer patterns of the filler statements. Table 3 presents an overview of the percentages of “true” answers to the filler statements.

As is clearly shown, cognitive load has no effect whatsoever on the answer pattern observed for the filler items. An ANOVA with cognitive load and working memory capacity as independent variables and the average number of “true” answers as dependent variable shows no overall effect of cognitive load (51% vs. 51%; $n = 92$), nor of working memory capacity ($n_{\text{lowest}} = 31$; $n_{\text{middle}} = 30$; $n_{\text{highest}} = 31$; 51%, 49%, and 53%).

Consistency analysis

We presented participants with 10 underinformative statements. In this section, we verify whether participants remain consistent over the trials. An overview of the consistency of the answer patterns is given in Table 4. Participants with 9 or 10 consistent answers were classified as consistent. All other participants were considered inconsistent.

Overall, we observed a high number of consistent responses (84%). Even if we adopt a stricter criterion—that is, including only participants who gave the same response for all 10 underinformative statements, the number of consistent responses remains fairly high (64%, of which 18% logical and 46%

Table 4. Overview of the consistency in answer patterns for the underinformative statements in Experiment 1 by cognitive load and number of trials

	Shift in consistency pattern from low to high load N = 92	Low cognitive load		High cognitive load	
		20 trials N = 49	30 trials N = 43	20 trials N = 49	30 trials N = 43
Consistent logic	+3	20	21	22	26
Consistent pragmatic	-5	69	58	59	58
Inconsistent	+2	10	21	18	16

Note: Results are expressed as the percentage of participants representing the answer pattern.

pragmatic). We did not observe an effect of cognitive load on the consistency of the individual answer patterns (Wilcoxon Matched Pairs Test, *ns*).

Discussion

The aim of this experiment was twofold. On the one hand, we intended to replicate the findings of De Neys and Schaeken (2007), who showed that imposing a cognitive load on working memory space resulted in a reduction in pragmatic answers. On the other hand, we aimed to test whether this effect was mediated by the interindividual differences in participants' working memory capacity, as suggested by Feeney et al. (2004). We measured the working memory capacity of the participants with the Operation Span Task for group testing. We operationalized the degree of cognitive load with a spatial memory task. This task was added to the statement evaluation task in a secondary task paradigm, allowing us to place either heavy or little load on cognitive resources. We observed an effect of this second task. Fewer pragmatic answers were given under high cognitive load than under low cognitive load. This result is in line with the main thesis of the contextualists, who claim that the pragmatic interpretation—in other words, the scalar implicature—arises at a larger cost and will therefore become less likely when cognitive resources are loaded.

We presented the participants with 10 underinformative statements, giving us the ability to perform an analysis of the individual consistency patterns of the participants. We observed highly consistent answer patterns. This strengthens

previous findings, since it ensures that the results can be attributed to the interpretation of the words and not to random fluctuations (see Oberauer et al., 2007, for an analysis on this topic).

We also measured the cognitive resources of the participants in a working memory capacity task. It is assumed that participants with lower working memory capacity can rely less on their cognitive resources than participants with high working memory capacity. The contextualists would therefore expect participants with a lower capacity to produce fewer pragmatic answers, particularly under cognitive load. Our experiment confirmed that there is an interaction between working memory capacity and cognitive load that affects the answer pattern. Participants with a low working memory capacity gave fewer pragmatic answers under cognitive load, while the answers of participants with more cognitive resources were not influenced by imposing cognitive load on their working memory. These results support the contextualist view on scalar implicatures. Our data are also partly in line with De Neys and Schaeken (2007), who observed a limited effect of cognitive load across all participants. The results of our experiment, however, do not confirm the positive correlation between working memory capacity and logical responses to underinformative statements, as observed by Feeney et al. (2004; the correlation between the raw GOSPAN scores and the number of logical answers under low and high cognitive load are $r = -.04$ and $r = .04$, respectively). No main effect of working memory capacity was observed.

The attentive reader may have noticed the unexpected U-shaped form of the accuracy data. Further

analysis did not reveal any differences between the three working memory capacity groups. Since there is no statistical evidence for the robustness of this observation, we assume that it is due to an (undesirable) coincidence. A more daring explanation, or rather speculation, for the observation of the relatively low percentage of pragmatic answers in the middle working memory capacity group is that they adopted a different task strategy. In the Operation Span Task this group showed a (nonsignificant) tendency to perform the task faster than the low and high working memory capacity group. This group of participants may have given priority to speed over accuracy in order to minimize the cognitive cost involved in the task. If they adopted the same task strategy in the statement evaluation task, the lower number of pragmatic answers can be explained by the contextualists' argumentation that a larger cognitive effort is required for a pragmatic answer. Again, this explanation is very tentative but may be inspiring for further research.

A possible criticism against the contextualist view may be the following: Although the results of the cognitive load manipulation support this view, the high percentage of pragmatic answers (68%), even in the (relatively) low working memory capacity group (73%), does not. The contextualist counterargument is that drawing scalar implicatures does not require a large cognitive effort, since they are so widespread in daily communication. The key point in the debate, however, is that implicatures are not completely automatic, which is confirmed by our data. Nevertheless, we obtain a high percentage of pragmatic answers, and, as is shown in Table 5, it is striking that the percentages of pragmatic answer patterns differ so widely between different studies. Therefore, it might be useful to take a closer look at the specifics of the task and answer patterns in previous studies on scalar <all, some> implicatures with adults. Of course, many subtle factors may account for this variation in results, such as discrepancies in instructions, item choice, number of trials, type of task, or language. One factor that we consider to be quite important and which has received little attention so far is language (but see

also Pouscoulous et al., 2007, on this topic). The meaning of "some" in English may not correspond entirely with its Dutch translation "sommige". For French, Pouscoulous et al. (2007, Experiment 3) has already demonstrated that children behaved differently in an action-based implicature task depending on whether "certains" or "quelques" (both are translations of "some") was used to describe the situation. This aspect certainly requires further study.

A second variable factor is the type of relation expressed in the statements. While Bott and Noveck (2004), among others, used <element, category> statements such as "some elephants are mammals", Guasti et al. (2005), among others, used <object, property> statements such as "some elephants have trunks". Intuitively, the former relation seems to be stricter than the latter, which may have an influence on the percentage of pragmatic answers and on the consistency in answer patterns. As shown in Table 5, Bott and Noveck (2004, Experiment 3) observed fewer consistent answer patterns (22%) than other studies that mainly employed <object, property> relations. In Experiment 1, we observed a substantially higher level of consistent answer patterns (64% if we use the same strict criterion). Nevertheless, a controlled study comparing both types of statements would be useful.

Finally, another factor that may influence the percentage of pragmatic answers is the number of trials included in the task. Table 5 shows that experiments with a high ratio of filler statements (0.83) lead to fewer consistent answer patterns (28%–74%) than do experiments with relatively smaller proportions of filler statements. Pouscoulous et al. (2007) argue that a higher number of controls, which typically avoid highlighting the (weak–strong) contrasts that are so critical for scalar inferences, leaves children unprepared to detect such contrasts when implicatures are potentially called for. De Neys and Schaeken (2007) argue that the relatively low number of filler statements in their experiment offered participants more opportunities to come to a pragmatic interpretation, as repetition helps to make a process more intuitive and cognitively less demanding. Vice versa, a larger proportion of filler

Table 5. Overview of answer patterns for underinformative statements in scalar <all, some> implicature studies with adults⁴

Experiment	Task type	Item type	Language	Number of trials	Types of filler items	Proportion of filler statements	Pragmatic answers (%)	Consistent (%)		
								Pragmatic	Logic (true)	Overall
Bott & Noveck (2004), Exp. 3	SET	E-C	French	54	5	.83	61	22	6	28
De Neys & Schaeken (2007)	SET	E-C	Dutch	20	5	.5	76	n.a.	n.a.	n.a.
Feeney et al. (2004), Exp. 3	SET	O-P	English	30	5	.83	23	6	51	57
Guasti et al. (2005), Exp. 1, adults	SET	O-P	Italian	30	5	.83	50	37	37	74
Guasti et al. (2005), Exp. 4, adults	TVJT	O-P	Italian	9	2	.44	83	83	17	100
Noveck (2001), Exp. 3, adults	SET	O-P	French	30	5	.83	41	33	40	73
Noveck & Posada (2003)	SET	O-P	French	75	2	.67	63	63	37	100
Experiment 1 (10 fillers)	SET	E-C	Dutch	20	5	.5	72	51	16	67
Experiment 1 (20 fillers)	SET	E-C	Dutch	30	5	.67	64	39	20	59
Experiment 2 (20 fillers)	SET	E-C	Dutch	30	5	.67	81	54	10	63
Experiment 2 (50 fillers)	SET	E-C	Dutch	60	5	.83	67	24	14	38

Note: We used a very strict criterion for consistency (in some studies the only available one). Therefore these percentages differ from those shown in Tables 4 and 6. SET = statement evaluation task. TVJT = truth value judgement task. E-C = element-category. O-P = object-property. n.a. = not available.

statements interrupts repetition, making it more difficult to develop a response strategy (and thus to generate a consistent answer pattern) for the underinformative statements. So, they both claim that when more distractor items are added, fewer resources are available to produce implicatures. Whether the effect of repetition on cognitive demands of the task is indirect, through predictability, as Pouscoulous et al. propose or direct as De Neys and Schaeken propose is left open for

discussion here. (See also: Botvinick, Braver, Barch, Carter, & Cohen, 2001; Shiffrin & Schneider, 1977, for the effect of repetition on cognitive demands.)

In Experiment 1, we also introduced two different proportions of filler statements. In one condition, we introduced 10 filler statements to keep the number of filler statements and underinformative statements equal. In another condition, we introduced 20 filler statements in order to

⁴ No consistency information is reported for Hurewitz, Papafragou, Gleitman, and Gelman (2006) and Pouscoulous et al. (2007) because their experimental task deviated considerably from the SET and in the case of Pouscoulous et al. (2007) because only one underinformative statement was presented. The experimental tasks of Breheny et al. (2006) and Bezuidenhout and Morris (2004) also deviated too much to be included, and, what is more, they only reported reaction times.

harmonize the clearly false, clearly true, and underinformative statements. Although it was not the intention in Experiment 1 to set up a manipulation of interruption of the underinformative statements, we may have another look at the results, given the discussion above. As is mentioned in the response analysis, no effect of this variable on the percentage of pragmatic answers is observed. We performed a new consistency analysis to determine whether the ratio of participants that answered consistently (to those who answered inconsistently) is higher in the group with 20 trials than in the group with 30 trials. This is clearly not the case (low load: 5/44 vs. 9/34; high load: 9/40 vs. 7/36; see also Table 4). Possibly the proportion of filler items was not large enough to disrupt the response strategy. In Experiment 2, we further study the effect of repetition of underinformative statements on the answer pattern.

EXPERIMENT 2

From Experiment 1 we learned that a (supposedly small) cognitive cost is involved in drawing a pragmatic implicature. Nevertheless, we observed a rather high percentage of pragmatic answers. This may be attributed to a number of factors. One of them is the relatively low proportion of filler statements included in the experiment. In Experiment 2 we manipulate this factor. One group of participants will receive 20 filler statements and 10 underinformative statements, resulting in a ratio of 0.67. Another group of participants will receive 50 filler statements and 10 underinformative statements, resulting in a ratio of 0.83. We predict that the introduction of more filler statements will lead to a decrease in consistency and in the overall percentage of pragmatic answers.

Method

Participants

The 158 second bachelor students of the University of Leuven who participated in the experiment took part to fulfil credit requirements.

Materials, design, and procedure

For the statement evaluation task, see Experiment 1. As in the previous experiment, the number of underinformative statements was 10. The number of filler statements was 20 or 50, and the five types of filler statements are equally represented.

The independent variable was the number of filler statements (20 vs. 50). The dependent variable was the type of response (logical–pragmatic).

The statement evaluation task is computer driven and completed collectively after random allocation of the participants to the two experiment conditions. Because of the large number of trials, we expected that a considerable share of the participants assigned to the 50-statements task would be excluded based on the 15% exclusion criterion (see Experiment 1), due to a lack of concentration. Therefore, we allocated the largest number of participants to this task (106 students), while only 52 students were assigned to the task with 20 filler statements.

Results

Exclusion criterion

Contrary to our expectations (see Procedure), no participants had to be excluded based on the exclusion criterion mentioned in the Materials section (see Experiment 1).

Response analysis

A *t* test shows that fewer pragmatic answers were given to the underinformative statements by the participants in the group with 50 filler statements (67%) than by the participants in the group with 20 filler statements (81%), $t(156) = 2.38$, $p < .05$.

Consistency analysis

Table 6 offers an overview of the consistency of the participants for the underinformative statements based on the number of filler statements. Participants with 9 or 10 consistent answers were classified as consistent. All other participants were considered inconsistent.

Overall, we observed that 70% of answer patterns were consistent. If we adopt a stricter criterion—that is, including only participants who gave the same response for all 10 underinformative

Table 6. Overview of the consistency in answer patterns for the underinformative statements in Experiment 2 by number of filler statements

	Filler statements	
	20 (<i>N</i> = 52)	50 (<i>N</i> = 106)
Consistent logic	12	18
Consistent pragmatic	71	45
Inconsistent	17	37

Note: In percentages.

statements—the percentage of consistent answer patterns drops to 46% (63% in the 20 filler statements group; 38% in the 50 filler statements group). It is clear that there is a correlation between the number of filler statements and response behaviour. The ratio of consistent answers (to inconsistent ones) in the 20 filler statements group is higher than that in the 50 filler statements group (43/9 vs. 67/39), $\chi^2(1) = 6.26$, $p < .05$. If we compare the number of pragmatic consistent answers alone against all other answers, the effect becomes even more pronounced (37/15 vs. 48/58), $\chi^2(1) = 9.39$, $p < .005$.

Discussion

By comparing methodological details of Experiment 1 with similar experiments in the literature, we observed that the proportion of filler statements varied extensively between the experiments. We also observed that this variation agreed with the variation in the consistency of the answer patterns. In Experiment 2, we experimentally tested whether extensively increasing the proportion of filler statements had an effect on the response behaviour. We predicted that participants who were confronted with more filler statements would be less likely to develop a consistent response strategy and that they would give fewer pragmatic answers. The rationale behind this prediction is that a constant repetition of underinformative statements makes the task easier since the predictability increases, and participants tend to develop a response strategy. With more filler statements

added, no response strategy can be developed, leaving the participant with fewer resources available to draw the implicature. That is exactly what the data confirmed.

Guasti et al. (2005) observed a strong bimodal distribution in their experiments. This led them to assume that participants have a strong tendency to develop response strategies, especially in the SET and to a lesser extent in the truth value judgement task (TVJT). We showed that a bimodal response pattern can be interrupted in the SET by increasing the number of filler items. Guasti et al. follow a defaultist line of reasoning and suggest that whatever answer participants give, they have first recognized that the pragmatic interpretation is false, so they have automatically made the pragmatic inference, but in the case of a logical answer they assume that in the experimental setting a logical answer is required, or they extend the context in a noncanonical way to make the logical answer the “true” answer. However, if the pragmatic answers were fully automatic, increasing the number of filler items should have no influence on the pragmatic answers or lead to a decrease in (costly) logical answers. The reverse was observed in Experiment 2.

It is also interesting to note that the drop in consistency with 50 filler statements (38% if the strict criterion is adopted) is perfectly in line with the results (22%) obtained by Bott and Noveck (2004, Experiment 3). Although they did not manipulate the number of filler statements to obtain this effect, the increase may have caused the equivocality in the participants’ responses, which provided the authors with within-subject data in a well-controlled task that showed that logical responses were faster than pragmatic ones. The observed results are also in line with the view of De Neys and Schaeken (2007) that an increase in the (proportion of) filler statements leads to fewer consistent answer patterns. In addition, they contradict an earlier suggestion made by Feeney et al. (2004, Experiment 3, p. 127) who claim that “one way to reduce participants’ use of strategies might be to include fewer trials overall and to increase the types of trial”. If this suggestion

were correct, we would have observed fewer consistent answer patterns for the underinformative statements in the condition with 20 filler statements than in the condition with 50 filler statements.

CONCLUSION

The present study contributes to the ongoing debate about the automaticity of scalar implicatures. A large number of studies on this topic focus on the scale <all, some> since it has a logical norm of correctness. The logically correct interpretation of “some A are B” is that “some and possibly all A are B”. However, this often conflicts with the intuitive idea (the implicature) that the weaker term “some” is used when the stronger term “all” is not applicable. The main argument in the debate for or against the automaticity of scalar implicatures is based on the cognitive cost related to drawing the implicature. Taking a defaultist stance, Levinson (2000) claims that the implicature is drawn automatically. Noveck and a number of other researchers who adopt a contextualist perspective (Chevalier, Noveck, Nazir, Bott, Lanzetti, & Sperber, 2008; Noveck, 2001) argue that drawing the implicature requires a (small) cognitive cost. Most experiments provide indirect evidence against the automaticity of scalar implicatures, mainly through developmental studies (Guasti et al., 2005; Noveck, 2001; Papafragou & Musolino, 2003; Pouscoulous et al., 2007) and through latency information (Bott & Noveck, 2004; Breheny et al., 2006; Feeney et al., 2004; Noveck & Posada, 2003). In the present study, we provided direct evidence against the automaticity of scalar implicatures through manipulation of the cognitive load imposed on the participants (see also Bott & Noveck, 2004, and De Neys & Schaeken, 2007, for direct evidence).

In Experiment 1, we manipulated the cognitive load by introducing a second, distracting task. Each participant performed the statement evaluation task twice, once combined with a task that required much of their cognitive resources, and once combined with a task that required less of their

cognitive resources. We also measured working memory capacity, as we expected the cognitive load manipulation to be mediated by the participants' working memory capacity. In line with the contextualist view, we predicted that the cognitive load manipulation would have more effect on the group with (relatively) lower working memory capacity. These expectations were confirmed. Participants gave fewer pragmatic answers under high cognitive load, but there was also a correlation with working memory capacity. After separate analysis, it was clear that the cognitive load effect only remained significant for the group with the lowest working memory capacity. In Experiment 2, we increased the number of filler statements, leaving the number of underinformative sentences constant. We observed that participants became less consistent in their answer patterns when the proportion of filler statements was high. This is an indication that it is hard for participants to develop a response strategy in that case. Moreover, we observed that increasing the number of filler statements has the same influence as the cognitive load manipulation in Experiment 1. The higher the number of filler statements, the fewer pragmatic answers are given. Note that if the implicature were drawn automatically, as Levinson (2000) predicted, no effect of proportion of filler statements should be observed. Both experiments provide further evidence in favour of the contextualist view that the pragmatic implicature is not automatically drawn, but requires some cognitive effort.

Original manuscript received 28 September 2009

Accepted revision received 29 April 2011

First published online 7 September 2011

REFERENCES

- Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, 49A, 5–28.
- Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 12–23.

- Bezuidenhout, A. L., & Morris, R. K. (2004). Implicature, relevance and default pragmatic inference. In I. A. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 257–282). Basingstoke, UK: Palgrave Macmillan.
- Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, 51(3), 437–457.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108, 624–652.
- Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, 100(3), 434–463.
- Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance theory: Applications and implications*. Amsterdam, The Netherlands: John Benjamins.
- Carston, R. (2004). Relevance theory and the saying/implicating distinction. In L. Horn & G. Ward (Eds.), *Handbook of pragmatics* (pp. 633–656). Oxford, UK: Blackwell.
- Chevalier, C., Noveck, I., Nazir, T., Bott, L., Lanzetti, V., & Sperber, D. (2008). Making disjunctions exclusive. *Quarterly Journal of Experimental Psychology*, 61(11), 1741–1760.
- Chierchia, G. (2004). Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond*. Oxford, UK: Oxford University Press.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, 19(4), 450–466.
- De Neys, W. (2006a). Automatic-heuristic and executive-analytic processing during reasoning: Chronometric and dual-task considerations. *Quarterly Journal of Experimental Psychology*, 59(6), 1070–1100.
- De Neys, W. (2006b). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, 17(5), 428–433.
- De Neys, W., d'Ydewalle, G., Schaeken, W., & Vos, G. (2003). A Dutch computerized and group administrable adaptation of the operation span test. *Psychologica Belgica*, 42, 117–190.
- De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology*, 54(2), 128–133.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1), 19–23.
- Feeney, A., Scafton, S., Duckworth, A., & Handley, S. J. (2004). The story of some: Everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology*, 58(2), 121–132.
- Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, 20(5), 667–696.
- Horn, L. R. (1972). *On the semantic properties of logical operators in English*, Unpublished doctoral dissertation, University of California, Los Angeles (Distributed by Indiana University Linguistics Club, 1976).
- Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Shiffrin (Ed.), *Meaning, form and use in context: Linguistic applications* (pp. 11–42). Washington, DC: Georgetown University Press.
- Horn, L. R. (2001). *A natural history of negation*. Chicago, IL: University of Chicago Press (Original work published 1989).
- Horn, L. R. (2004). Implicature. In L. R. Horn & G. L. Ward (Eds.), *Handbook of pragmatics* (pp. 3–28). Oxford, UK: Blackwell.
- Hurewitz, F., Papafragou, A., Gleitman, L., & Gelman, R. (2006). Asymmetries in the acquisition of numbers and quantifiers. *Language, Learning and Development*, 2(2), 77–96.
- La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16, 1118–1133.
- Levinson, S. (1983). *Pragmatics*. Cambridge, UK: Cambridge University Press.
- Levinson, S. (2000). *Presumptive meanings: The theory of generalised conversational implicature*. Cambridge, MA: MIT Press.
- Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, 130, 621–640.
- Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, 78(2), 165–188.

- Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, 85(2), 203–210.
- Noveck, I. A., & Reboul, A. (2008). Experimental pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences*, 12(11), 425–431.
- Oberauer, K., Geiger, S. M., Fischer, K., & Weidenfeld, A. (2007). Two meanings of “if”? Individual differences in the interpretation of conditionals. *Quarterly Journal of Experimental Psychology*, 60(6), 790–819.
- Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition*, 86(3), 253–282.
- Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition: A Journal of Developmental Linguistics*, 14(4), 347–375.
- Rips, L. J. (1975). Quantification and semantic memory. *Cognitive Psychology*, 7, 307–340.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127–190.
- Smith, C. L. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology*, 30, 191–205.
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.
- Wilson, D., & Sperber, D. (2004). Relevance theory. In L. Horn & G. Ward (Eds.), *Handbook of pragmatics* (pp. 607–632). Oxford, UK: Blackwell.