When People Are More Logical Under Cognitive Load

Dual Task Impact on Scalar Implicature

Wim De Neys and Walter Schaeken

University of Leuven, Belgium

Abstract. The present study introduces dual task methodology to test opposing psychological processing predictions concerning the nature of implicatures in pragmatic theories. Implicatures routinely arise in human communication when hearers interpret utterances pragmatically and go beyond the logical meaning of the terms. The neo-Gricean view (e.g., Levinson, 2000) assumes that implicatures are generated automatically whereas relevance theory (Sperber & Wilson, 1986/1995) assumes that implicatures are effortful and not automatic. Participants were presented a sentence verification task with underinformative sentences that have the potential to produce scalar implicatures like *Some oaks are trees*. Depending on the nature of the interpretation of *Some* (logical or pragmatic) the sentence is judged true or false. Executive cognitive resources were experimentally burdened by the concurrent memorization of complex dot patterns during the interpretation process. Results showed that participants made more logical and fewer pragmatic interpretations under load. Findings provide direct support for the relevance theory view.

Keywords: implicatures, pragmatics, executive processing

The complexity of human communication has been characterized as one of the hallmarks of our species. A striking demonstration of the sophisticated nature of our communication system is our ability to draw pragmatic inferences or implicatures. Often speakers intend to convey far more than the logical meaning of the words they utter and hearers readily retrieve the intended interpretation. Imagine, for example, that seven-time Tour de France winner Lance Armstrong states in an interview:

(a) Some cyclists use doping.

You would readily assume that Lance intends to say:

(b) Not all cyclists use doping.

Rather then:

(c) All cyclists use doping.

However, according to standard semantic accounts, *some* means *some and possibly all*. From a strictly logical point of view, Lance's utterance is, therefore, compatible with (c) and not with (b). In statement (b), *some* is interpreted pragmatically. Based on expectations about how a conversational exchange occurs the hearer goes beyond the literal, logical meaning of the uttered terms (Grice, 1989). Such pragmatic inferences or implicatures have been extensively studied in philosophical and linguistic works on verbal communication. The inference in the example has been dubbed *scalar implicature* since the constituting terms can be ordered on a scale of informativeness. The implicature arises when a less informative term (e.g., *some*) is taken to imply the denial of the more informative term (e.g., *all*). The scalar implicature is considered as the paradigmatic case for the study of implicatures.

There is an interesting debate in the literature concerning the way scalar inferences actually work (i.e., how a scalar inference manifests itself in real time). The dispute centers around the automaticity of the scalar implicature (see Bott & Noveck, 2004; Noveck, 2001; Noveck & Posada, 2003; Papafragou & Musolino, 2003). One school of thought, referred to as the *neo-Gricean* account (e.g., Levinson, 2000) claims that the pragmatic interpretation is actually the default interpretation in a concrete, communicative setting. It is assumed that every time *some* is encountered the hearer will automatically make the scalar implicature. Levinson, for example, has argued that scalar implicatures result from a Q-heuristic dictating that "What isn't said isn't." Therefore, whenever a listener hears a weak scalar term like *Some*, the listener will automatically assume that the speaker intended that a stronger term (i.e., *all*) is not warranted. The Q-heuristic dictates that if it is not said that *all* is the case, it is not. If the speaker intended *some* to imply the stronger term *all*, he or she should have said it explicitly. Hence, the hearer will always start by making the scalar implicature. An eventual logical interpretation of *some* can only arise in a later stage where the implicature is undone by the context.

A second approach, *relevance theory* (Carston, 1999; Sperber & Wilson, 1986/1995), assumes implicatures are

DOI 10.1027/1618-3169.54.2.128

not made by default. Relevance theory states that how far a hearer goes in processing an utterance's meaning is governed by principles concerning effect and effort. Listeners try to gain as many effects as possible for the least effort. According to relevance theory the logical interpretation of *some* (i.e., some and possibly all) could very well lead to a satisfying interpretation in an utterance. It is possible that the hearer will derive the scalar implicature and move to a pragmatic interpretation (i.e., some but not all) to make the utterance more informative but this enrichment is not automatic and will come at the cost of additional processing effort. Thus, from the perspective of relevance theory scalar implicatures are considered as effortful, nonnecessary inferences, whereas according to a neo-Gricean view it is precisely the occasional undoing of an automatic, default implicature that takes extra effort (for more discussion, see Bott & Noveck, 2004).

Intuitively, the neo-Gricean account seems to be the most plausible. If one looks at Lance's utterance (a), for example, it certainly *feels* like we infer (b) directly while the logical interpretation (c) seems to require far more active thinking. However, contrary to the intuitive appeal, recent experimental findings seem to provide some support for the relevance theory view. Noveck (2001), for example, showed that children made the logical interpretation of *some* more than adults (see also Papafragou & Musolino, 2003). Thus, the increase in pragmatic inference-making appears to be linked with greater access to cognitive resources (which one would naturally assume increases with age). In a series of latency studies, Bott and Noveck (2004, see also Noveck & Posada, 2003) observed that people needed more time for the pragmatic interpretation than for the logical one. They also found that limiting the time available for responding boosted the rate of logical interpretations. These findings suggest that making scalar implicatures is associated with correlates of effortful processing.

The present study introduced a dual task approach to settle the debate with a more direct test of the automaticity claim. Cognitive resources were experimentally burdened by imposing a resource-demanding secondary task during sentence interpretation. Participants were presented with so-called underinformative sentences like "some tuna are fish" and were asked to judge whether the sentence was true or false (e.g., Bott & Noveck, 2004). Since, by definition, all tuna are fish, the sentence will be judged false if one draws the scalar implicature and interprets the sentence pragmatically (e.g., some but not all tuna are fish). If one interprets *some* logically (e.g., some and possibly all tuna are fish) the sentence will be judged true.

If the scalar implicature is not automatic but requires effortful processing, making the implicature should be harder when cognitive resources are burdened. Therefore, from the perspective of relevance theory one predicts that the rate of pragmatic inferences will decrease under cognitive load. On the other hand, if one believes that the implicature is made automatically and it is the logical interpretation that requires additional processing, burdening the cognitive resources should hamper the logical interpretation process. Hence, people should be more likely to stick to the default pragmatic interpretation under load. From a neo-Gricean view one therefore predicts that the cognitive load will boost the rate of pragmatic interpretations.

Notice that pragmatic theorists and previous experimental studies have not characterized the exact nature of the alleged effortful processing. The present study focused on the role of executive working memory resources since these are widely recognized as the quintessential component of human cognitive capacity (e.g., Engle, Tuholski, Laughlin, & Conway, 1999).

Experiment

Participants verified sentences while they concurrently tried to remember a briefly presented visual dot pattern (e.g., Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). The complexity of the dot pattern was manipulated so that storage of the pattern in a control condition would be less demanding.

Method

Participants

A total of 56 first-year psychology students of the University of Leuven participated in return for course credit.

Material

Sentence Verification Task

Participants provided true/false judgments for 10 underinformative sentences presented on a computer screen in front of them. Content of the sentences referred to categories and exemplars (e.g., Some "*exemplar*" are "*category*"). As in the study by Bott and Noveck (2004, Experiment 3), we did not impose any specific interpretation. The only instructions participants were given was to respond "true" if they thought the sentence on the screen was true, or "false" if they believed the sentence to be false. Participants typed down their response using the numpad ("1" – True or "2" – False).

Participants also judged 10 filler trials where patently true (e.g., Some birds are eagles) and patently false (e.g., Some tigers are fish) sentences were presented. The underinformative and filler sentences were presented in random order. Participants were presented two verification task sets (with 10 underinformative and 10 filler sentences in each set, see Appendix), one while memorizing easy control patterns and one while memorizing complex load patterns. Thus, the load factor was manipulated within-subjects.

Both sentence sets were used with equal frequency in the load and control condition.

Dot Memory Task

The dot memory task is a classic spatial storage task (e.g., Bethell-Fox & Shepard, 1988; Miyake et al., 2001). For the present study a 3 × 3 matrix filled with three to four dots was presented for 850 ms. Participants memorized the pattern and were asked to reproduce it afterwards.

In the load trials the matrix was filled with a complex four-dot pattern (i.e., a "two- or three-piece" pattern based on the work of Bethell-Fox & Shepard, 1988, and Verschueren, Schaeken, & d'Ydewalle, 2004, see Figure 1). Miyake et al. (2001) established that storage of similar complex dot patterns tapped executive resources.

In the control trials the pattern consisted of three dots on a horizontal line (i.e., a "one-piece" pattern in Bethell-Fox & Shepard's terms). This simple and systematic pattern should only minimally burden the executive resources (e.g., De Neys, 2006; Miyake et al., 2001).

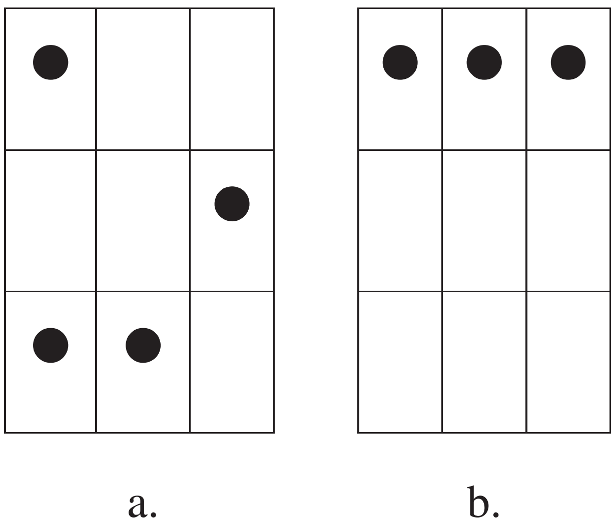

Figure 1. Examples of the dot patterns in the load (a) and control trials (b).

Procedure

Participants were tested in small groups. It was randomly determined which sentence set and which load type each participant started with. All sessions started with a demonstration of the storage task. On two practice storage items (one with a simple and one with a complex pattern) an empty response matrix was presented 1 s after the pattern had been presented. Participants used the keypad to indicate the location of the dots. Instructions stressed that it was crucial that the dot patterns were reproduced correctly in the upcoming task.

A verification task trial started with a brief presentation of the dot pattern for 850 ms. Next, the sentence was presented and remained on the screen until the participant made a response. Afterwards, an empty matrix was presented and participants had to reproduce the dot pattern. Participants received feedback on whether the pattern had been reproduced correctly and were reminded that they had to remember the complete pattern correctly. The procedure was clarified with two practice sentences.

Results and Discussion

Dot Memory Task

Results for the dot memory task indicated that the task was properly performed. Recall performance was overall high and showed little interindividual and intraindividual variation. The mean number of correctly localized dots for the complex four-dot patterns was 3.75 (SD = .25) and 2.90 (SD = .17) for the simple three-dot patterns. Thus, overall, about 94% of the complex and 97% of the simple patterns were reproduced correctly. Moreover, dot recall performance and performance on the sentence verification task were not associated. The correlation between the mean number of pragmatic responses and mean number of correctly localized dots was small and not significant (r = .06, p > .6). These data clearly show that participants were not trading-off performance in the two tasks.

Sentence Verification Task

Underinformative Sentences

Remember that when underinformative sentences are interpreted pragmatically they will be judged "false" in the sentence verification task. Overall, participants gave a mean number of 7.61 "false" responses to the underinformative sentences. Thus, the vast majority (about 76%) of participants' interpretations were pragmatic in nature. The crucial finding is that participants made significantly fewer pragmatic interpretations when they had to memorize the demanding complex patterns (M = 7.32, SD = 3.98) than when memorizing the easy control patterns (M = 7.89, SD = 3.68), $t(55)$ = 2.18, p < .04, d = .41. This trend is consistent with the prediction from a relevance theory perspective. Concurrent memorization of the complex dot pattern decreases rather than boosts the tendency to make pragmatic interpretations.

Filler Sentences

To check whether the load did not interfere with basic sentence comprehension we calculated the mean number of correct responses to the filler sentences. Performance was uniformly high. Participants responded correctly to the filler sentences both when memorizing the control (M = 9.36, SD = .77) and complex patterns (M = 9.3, SD = .1.01), $t(55)$ = .37, p = .72. The high overall rate of correct filler responses (less than 7% errors) establishes that participants were not simply guessing or superficially accepting under secondary task load. In principle, one could suggest that the secondary task load triggers a general tendency to select "true" responses. Hence, the load effects on the underinformative sentences would not point to a difficulty to derive the implicature *per se*. Note, however, that 4 out of 10 filler sentences were patently true. Hence, if participants would simply accept or guess under load they should err on 50% to 60% of the filler trials. The high rate of correct responses on the filler items establishes that the cognitive burden specifically interfered with drawing scalar implicatures.

Sentence Verification Latencies

Participants' response latencies (i.e., the time between sentence presentation and the key press) in the verification task were also recorded. We mentioned that Bott and Noveck (2004) found that making pragmatic interpretations took more time than making logical interpretations. Because of the low number of logical interpretations in the present study, a direct comparison between latencies for pragmatic and logical interpretations was not very informative here[1]. However, we could compare the impact of cognitive load on the latencies for pragmatic interpretations *per se*. This allows an additional validation of our findings. If making pragmatic interpretations is cognitively demanding than one would predict that on those occasions where people do manage to make pragmatic interpretations under complex load this would come at the cost of additional processing time compared to the control condition. This prediction was confirmed. Pragmatic interpretations under load (M = 6727 ms, SD = 1394) took about 700 ms longer than pragmatic interpretations on control trials (M = 6031 ms, SD = 1155), $t(45) = 2.73$, $p < .01$, $d = .57$. For completeness, we also looked at the load impact on the latencies for logical responses. Interestingly, consistent with the idea that logical interpretations are made automatically, the latencies did not differ on load (M = 6388 ms, SD = 1394) and control (M = 6431, SD = 1176) trials, $t(16) = -.06$, $p > .9$. While the nonsignificant effect for the logical interpretation might be attributed to the limited reliability of the latency estimation for the infrequent logical responses, findings were further supported by examining latencies for the filler items. Filler items are consistently solved correctly and are also assumed to be processed automatically. Results indeed showed that filler latencies were not affected by load. Verifying filler sentences took about the same time under load (M = 6236, SD = 1394) and control (M = 5997, SD = 1036), $t(55) = 1.438$, $p > .15$. This indicates that the increased latencies for pragmatic interpretations under complex load are specifically tied to the demanding nature of the pragmatic processing.

General Discussion

The present study introduced dual task methodology to test opposing processing assumptions in pragmatic theories concerning the nature of scalar implicature. The intuitively appealing neo-Gricean account considers implicatures as automatic, default inferences that will need to be overridden to arrive at a logical interpretation. From the perspective of relevance theory the logical interpretation is more basic and implicatures are considered additional, nonautomatic, cognitively demanding inferences. The present findings clearly supported the relevance theory view of scalar implicature. Burdening participants' cognitive resources with the memorization of the complex dot patterns decreased the rate of pragmatic inferences. Hence, contrary to the neo-Gricean account, people became more "logical" under cognitive load. If it were the logical interpretation that required additional effortful processing, one would have expected to see more pragmatic and fewer logical interpretations under complex load. The decreased pragmatic interpretations directly establish that making scalar implicatures is not automatic but requires effortful, cognitive processing.

One aspect of the present study that deserves some further comment is the high (76%) overall percentage of pragmatic interpretations. In previous scalar implicature studies pragmatic and logical interpretations were typically made with roughly equal frequency (e.g., Experiment 3 in Bott & Noveck, 2004, Noveck & Posada, 2003). One difference between these studies and the present one is the ratio of underinformative and filler sentences. Bott and Noveck, for example, presented participants 54 sentences (5 out of 6 were fillers). Because of the demanding nature of the cognitive load manipulation (and the repeated testing) such a large number of trials would have been inappropriate in the present experiment. To keep participants optimally concentrated we decided to stick to 20 trials (1 out of 2 were fillers). However, relatively speaking, this implies that deriving the pragmatic interpretation will be much more frequently required in our study (i.e., every second trial). This might have made the pragmatic processing somewhat less demanding. It is well established that a more continuous repetition of a process can help automatize a task and reduce the need for cognitive control (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Shiffrin & Schneider, 1977). Studies on the classic Stroop task, for example, have typically observed that difficulties on the crucial incongruent trials (e.g., reading the word "red" when it is printed in yellow ink) are reduced when these trials become more frequent (e.g., Lindsay & Jacoby, 1994). In this sense one might argue that our verification task is somewhat easier than the version adopted by Noveck and colleagues. This would explain the higher rate of pragmatic interpretations.

1 There were only seven participants for whom we could directly compare response times for pragmatic and logical interpretations under control and complex load. None of the findings reached significance.

Manipulating the relative frequency of the underinformative sentences might prove to be an interesting new tool to examine the manifestation of scalar implicatures.

It could be argued that the impact of our cognitive load manipulation was not spectacular. Indeed, the complex dot memorization only decreased the rate of pragmatic interpretations by about 10% and participants still made about 70% pragmatic interpretations under load. However, it should be clear that the size of the load effect is not the crucial issue here. First, one could always try to obtain stronger effects by imposing more demanding secondary tasks (e.g., increasing the number of dots) or, as stated above, by making the verification task itself more demanding (e.g., possibly by decreasing the relative frequency of underinformative sentences). Second, and more important, it is obvious that no one in the relevance theory camp would claim that drawing an implicature puts a *massive* burden on our cognitive resources. Given the prevalence of pragmatic inferences in daily life, it would, indeed, be hard to see how people would manage to communicate if every implicature involved a major cognitive cost. Thus, it is rather likely that implicatures require a relatively small cognitive involvement. The important point is that they are not completely automatic. The effect might be small but what matters is that the effect is there. The basic finding that a cognitive burden results in more logical and fewer pragmatic interpretations directly contradicts the prediction from a neo-Gricean account of scalar implicature.

We stated that pragmatic theorists have not yet characterized the exact nature of the alleged effortful, cognitive processing. The memorization of complex dot-patterns has been shown to specifically burden the executive component of the human working memory system (Miyake et al., 2001). Hence, the present findings indicate that the effortful processing that is required in deriving scalar implicatures specifically draws on these executive working memory resources. This stipulation should stimulate pragmatic theorists to link their work more closely with the rich psychological research tradition on the role of executive resources in higher-order cognition (e.g., Kane & Engle, 2002). We believe that such an approach will be especially fruitful to arrive at more fine-grained future characterizations of the cognitive operations underlying pragmatic inferences.

We finally note an interesting implication of the present findings for the field of deductive reasoning. It is well-established that fallacious deductive inferences can be often attributed to pragmatic interpretations of the premises (e.g., Begg & Harris, 1982; Braine & O'Brien, 1998; Johnson-Laird & Byrne, 2002). Influential dual-process theories of reasoning (e.g., Evans, 2003; Stanovich & West, 2000) have attributed this pragmatic modulation to the operation of a heuristic system that biases the operation of a second, analytic system mediating normative correct reasoning. The basic assumption is that the heuristic system operates automatically whereas the second system would be cognitively demanding and draw on executive working memory resources. The present study indicates that contrary to the popular conceptualization, pragmatic modulation is not a purely automatic process. People already need executive resources to derive the potentially biasing scalar implicatures. Consistent with recent findings in the reasoning field (e.g., De Neys, Schaeken, & d'Ydewalle, 2005a, 2005b), this implies that the basic characterization of an automatically operating heuristic-pragmatic reasoning system can be questioned. Our findings clearly established that the paradigmatic case for the study of pragmatics, the scalar implicature, is not made automatically but involves effortful, executive processing.

References

Begg, I., & Harris, G. (1982). On the interpretation of syllogisms. *Journal of Verbal Learning and Verbal Behavior, 21,* 595–620.

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.

Bott, L., & Noveck, I.A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language, 51*, 437–457.

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.

Braine, M.D.S., & O'Brien, D.P. (Eds.). (1998). *Mental logic*. Mahwah, NJ: Erlbaum.

Carston, R. (1999). Informativeness, relevance, and scalar implicature. In R. Carston, & S. Uchida (Eds.), *Relevance theory: Applications and implications*. Amsterdam: John Benjamins.

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17,* 428–433.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005a). Working memory and counterexample retrieval for causal conditionals. *Thinking and Reasoning, 11*, 123–150.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005b). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking and Reasoning, 11,* 349–381.

Engle, R.W., Tuholski, S.W., Laughlin, J.E., & Conway, A.R.A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*, 309–331.

Evans, J. St. B.T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences, 7,* 454–459.

Grice, H.P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Johnson-Laird, P.N., & Byrne, R.M.J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109*, 646–678.

Kane, M.J., & Engle, R.W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual differences perspective. *Psychonomic Bulletin and Review, 9*, 637–671.

Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Lindsay, D.S., & Jacoby, L.L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 219–234.

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*, 165–188.

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

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition, 86*, 253–282.

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.

Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and cognition*. Oxford: Basil Blackwell.

Stanovich, K.E., & West, R.F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences, 23*, 645–726.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2004). Everyday conditional reasoning with working memory preload. In *Proceedings of the Twenty-Sixth Annual Meeting of the Cognitive Science Society* (pp. 1399–1404). Mahwah, NJ: Erlbaum.

Received March 3, 2006
Revision received June 9, 2006
Accepted June 12, 2006

Wim De Neys

Postdoctoraal Onderzoeker FWO
Lab Experimentele Psychologie
K.U. Leuven
Tiensestraat 102
B-3000 Leuven
Belgium
Tel. +32 16 326143
Fax +32 16 326099
E-mail Wim.Deneys@psy.kuleuven.ac.be

Appendix

Table A1. Sentences used in the verification tasks

Underinformative	Filler
Set 1	
Some eels are fish.	Some birds are magpies. (true)
Some carp are fish.	Some insects are wasps. (true)
Some oaks are trees.	Some pigeons are insects. (false)
Some beeches are trees.	Some beetles are flowers. (false)
Some sparrows are birds.	All chrysanthemums are flowers. (true)
Some robins are birds.	All hazels are trees. (true)
Some flies are insects.	All trees are elms. (false)
Some mosquitoes are insects.	All fish are herrings. (false)
Some roses are flowers.	All daffodils are trees. (false)
Some tulips are flowers.	All sycamores are fish. (false)
Set 2	
Some ants are insects.	Some flowers are carnations. (true)
Some bees are insects.	Some trees are willows. (true)
Some canaries are birds.	Some crocuses are trees. (false)
Some blackbirds are birds.	Some poplars are fish. (false)
Some daisies are flowers.	All cod are fish. (true)
Some lilies are flowers.	All parrots are birds. (true)
Some firs are trees.	All birds are crows. (false)
Some birches are trees.	All insects are worms. (false)
Some trout are fish.	All pike are birds. (false)
Some sharks are fish.	All swallows are insects. (false)

Note. Sentences were translated from Dutch.

