Contribution of working memory in parity and proportional judgments

Jakub Szymanik and Marcin Zajenkowski
Institute of Artificial Intelligence, University of Groningen /
Faculty of Psychology, University of Warsaw

This paper presents experimental evidence on the differences in a sentence–picture verification task under additional memory load between parity and proportional quantifiers. We asked subjects to memorize strings of four or six digits, then to decide whether a quantified sentence was true for a given picture, and finally to recall the initially given string of numbers. The results show that: (a) proportional quantifiers are more difficult than parity quantifiers with respect to reaction time and accuracy; (b) maintaining either four or six elements in working memory has the same effect on the processing of parity quantifiers; (c) however, in the case of proportional quantifiers subjects perform better in the verification tasks under the six-digit load condition, and (d) even though the strings of four numbers were better recalled by subjects after judging parity there is no difference between quantifiers in the case of the six-element condition. We briefly outline two alternative explanations for the observed phenomena rooted in the computational model of quantifier verification and the different theories of working memory.

1. Introduction

In this paper we examine the contribution of working memory in the sentence–verification of natural language quantifiers. Psychological studies on information comparison from linguistic and pictorial sources have a long tradition (see e.g. Clark & Chase 1972; Moxey & Sanford 1993). However, only recently has research been initiated that bases psychological predictions on a computational model of quantifier verification proposed by linguists and logicians (see Szymanik 2007; Szymanik & Zajenkowski 2010a). In particular, this paper extends the results of Szymanik and Zajenkowski (2010b) by studying further the differences between

Belgian Journal of Linguistics 25 (2011), **176–194.** DOI 10.1075/bjl.25.08szy
ISSN 0774-5141 / E-ISSN 1569-9676

parity (divisibility) and proportional judgments. The first type of quantifiers say something about the parity of a set satisfying some condition, e.g.,

(1) An even number of referees reviewed that paper.
An odd number of reviewers enjoyed their job.

The second type expresses the proportion of elements possessing the property to all other elements, e.g.,

(2) More than half of the papers for this volume were rejected.
One third of all the authors come from Europe.

It is well known that neither class of quantifiers is definable in first-order (elementary) logic (see e.g. Peters & Westerståhl 2006). However, they differ a lot from the computational perspective. While the first class is recognizable by finite-state machines, the second one needs much more powerful devices with unbounded working memory, such as push-down automata (van Benthem 1986). Recently, there has been some controversy whether this difference has some psychological plausibility. In their neuropsychological experiment, McMillan et al. (2005) classified the parity and proportional groups as belonging to the same class of quantifiers not definable in elementary logic but Szymanik (2007) questioned that division and suggested that in fact they should be distinguished based on computational analysis.[1] In this paper we experimentally study the working memory resources needed to verify the two groups of quantifiers and show that indeed there are significant differences.

A linguist may worry at that point about our choice of quantifiers. Arguably, neither proportional nor parity quantifiers are the most usual type of quantification used in everyday language. However, subjects' accuracy in the verification tasks involving such sentences is usually at the ceiling. Therefore, those quantifiers can be naturally used in a verification study. Our goal is to show that computational complexity interacts with cognitive difficulty. We take into account the complexity of verification procedures (model-checking) and compare it to the behavioral indicators of subjects' performance in verification tasks. As a result we can conclude that the computational complexity of model-checking in the case of simple (monadic) quantifiers gives a nice prediction of the cognitive difficulties subjects face in the corresponding experimental tasks.

In our analysis we refer both to classical and modern approaches in psychological modelling of working memory. We focus on two questions:

- How does the processing of parity and proportional quantifiers relate to working memory?
- What are the elements of working memory responsible for the relationship?

This paper is structured as follows. First, we give a brief overview of models of working memory and quantifier verification. Next, we discuss our experiments. We conclude with a general discussion and a plan for future work.

1.1 Models of working memory

The concept of working memory was developed by Baddeley et al. (Baddeley & Hitch 1974; Baddeley 1986). These authors proposed an extension of the notion of short-term memory, suggesting that it should include three separable components: working memory consists not only of the two temporary storage units, phonological and visual, but also of a central executive, which takes the role of an attentional control system. Working together, these components form a unified memory system responsible for the performance of complex tasks.

Therefore, in the classical view, working memory involves two functions: processing and storage. The best-known tool to assess the psychological plausibility of the distinction is the reading span test (Daneman & Carpenter 1980). In the task, subjects read series of sentences and are asked to comprehend them as well as memorize the final word of each sentence. There are many variants of this test. For instance, Turner and Engle (1989) proposed an operation span task in which participants are required to solve a series of mathematical operations while trying to remember a set of unrelated words (see also for more details Unsworth, Heitz, Schrock & Engle 2005). All forms of the span tests require keeping in mind some data while processing other information. In experimental studies, a trade-off between storage and processing is usually observed (see e.g. Daneman & Carpenter 1980; Duff & Logie 2001). This means that the more information must be maintained in memory, the slower and poorer the concurrent processes become, and vice versa, with a higher demand on concurrent processing memory.

In more recent approaches to working memory modeling, the role of attentional processes is emphasized. For instance, Cowan (1995) argues that the core of working memory is the focus of attention in which a limited capacity of about four elements can be held and processed. The assignment of attention to the specific information makes this content highly accessible. This mechanism creates some sort of working space, which is the nature of working memory.

Oberauer et al. agree with Cowan that attentional focus plays the central role in working memory (see Oberauer, Suess, Wilhelm & Wittman 2003, 2008). Additionally, they elaborate on the model, distinguishing three cognitive functions. The first one is simultaneous storage and processing. The authors notice that, in the past, working memory was often defined only by this function (Oberauer, Suess, Wilhelm & Wittman 2003). They find that both terms, "storage" and "processing" were often used to refer to many unrelated phenomena. Hence, they propose to carefully separate the two concepts and as a result to understand processing as the transformation of information, while storage is the retention of briefly presented new information over a period of time when the information carrier is no longer present (see Oberauer, Suess, Wilhelm & Wittman 2003). The second function is the coordination of information elements into structures (more recently named *relational integration* by Oberauer, Suess, Wilhelm & Wittman 2008). This refers to the ability of building new relations between elements and integrating relations into structures. The third, and last, function is supervision: the control of ongoing cognitive processes such as the selective activation of relevant representations and the prevention of distraction (see Oberauer, Suess, Wilhelm & Wittman 2003, 2008). According to Suess et al. (2002), coordination and supervision are parts ofthe executive functions.

The other widely known theory which highlights the meaning of attention in working memory was proposed by Engle et al. (1999). They identified an executive mechanism, controlled attention, that selects which representations should remain suitably activated for the situational requirements. Its role is especially important in very demanding conditions such as working memory tasks. This situation usually involves interference, distraction, competition of activated representations for selection for action, and finally suppressing or inhibiting information irrelevant to the task.

Many studies have been devoted to understanding the relationship between working memory and language comprehension. Just and Carpenter (1992) even argue that the role of working memory in language comprehension is especially important because "comprehension entails processing a sequence of symbols that is produced and perceived over time." Working memory stores the representations emerging from processing a stream of successive written or spoken words. Moreover, it seems that the results of the span test are good predictors of language comprehension (see e.g. Conway & Engle 1996; Daneman & Green 1986; Just & Carpenter 1992; King & Just 1991) and other language-processing tasks, including quantifier verification (see Szymanik & Zajenkowski 2010b).

In the main study presented in this paper, we created a test which is a variation on the span task. We asked subjects to verify sentences while trying to remember strings of digits. We believe that this situation can reveal the specific contribution of working memory in parity and proportional judgments. As we argue below, the additional memory load presumably should influence only the latter type of judgments. Moreover, we were interested in which aspect, or aspects, of working memory are involved in processing a particular quantifier. Is it the storage functions or the executive processes that are responsible for the differences in the verification of parity and proportional quantifiers? Previous investigation in this area didn't provide an answer (McMillan, Clark, Moore, Devita & Grossman 2005).

1.2 Computational model of quantifier verification

The computational model of quantifier verification has been proposed by van Benthem (1986). He noticed that in order to verify first-order quantifiers we only need computability models which do not use any form of internal memory (data storage). Intuitively, to check whether sentence (3) is true, we do not have to involve short-term memory (working memory capacity).

(3) Every sentence in this paper is grammatically correct.

It suffices to read the sentences from this article one by one starting from the beginning. If we find an incorrect one, then we know that the statement is false. Otherwise, if we read the entire paper without finding any incorrect sentence, then statement (3) is true. We can proceed in a similar way for other first-order quantifiers. This procedure can be formalized in terms of finite automata, as illustrated by the automaton in Figure 1.

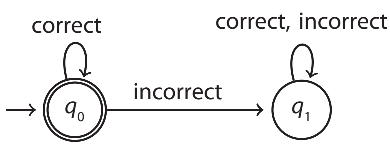

Figure 1. This finite automaton checks whether every sentence in the text is grammatically correct. It inspects the text sentence by sentence starting in the accepting state (double circled), q_0. As long as it does not find an incorrect sentence it stays in the accepting state. If it finds an incorrect sentence, then it already "knows" that the claim in sentence (3) is false, and moves to the rejecting state, q_1, where it stays no matter what sentence comes next.

How about non-elementary parity and proportional quantifiers? Parity quantifiers can still be recognized by finite automata. For example, consider sentence (4).

(4) An even number of the sentences in this paper are false.

When you find a false sentence you write "1" on the blackboard, if you find another one you erase "1" and put "0" again, then if you see another false sentence you put "1" in place of "0," and so on. At every moment you have only one digit on the blackboard no matter how long the paper is. This procedure can be implemented with the following finite automaton from Figure 2.

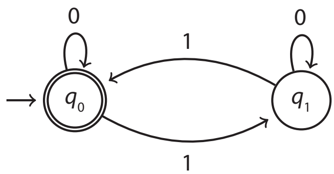

Figure 2. Finite automaton checking whether the number of "1"s is even.

However, for verifying proportional quantifiers we need computability models essentially stronger than finite automata. To compute the problem they have to make use of unbounded internal memory. Intuitively, to check whether sentence (5) is true we must identify the number of correct sentences and hold it in working memory to compare it with the number of incorrect sentences.

(5) Most of the sentences in this paper are grammatically correct.

Mathematically speaking, such an algorithm can be realized by a push-down automaton, PDA (see e.g. Hopcroft, Motwani & Ullman 2000), but it cannot be computed by any finite automaton.

This model leads to the psychological hypothesis that the cognitive difficulty of quantifier processing might be assessed on the basis of the complexity of the minimal corresponding automaton, as suggested by Szymanik (2007, 2009). Taking this perspective into account, Szymanik and Zajenkowski (2010a) conducted a reaction time experiment comparing various classes of quantifiers and showed that increased reaction time can be explained by the minimal automaton corresponding to the quantifier. Among others, the results indicate that quantifiers recognizable by finite automata are processed faster than proportional quantifiers. This is consistent with the computational analysis. Therefore, there is not only a quantitative but also a qualitative difference between the memory resources needed to compute these two types of quantifiers. This conclusion also follows from the differences in the brain recruitments observed by McMillan et al. (2005).

In the follow-up study, Szymanik and Zajenkowski (2010b) implemented a quantifier verification task under additional memory loads. The experiment was constructed in an analogous way to a span task: subjects were asked to memorize sequences of digits containing four or six elements, next they had to solve a quantifier verification task, and finally recall the digits. The data obtained revealed that in the four-element load condition, the most difficult were the proportional quantifiers "more than half" and "less than half" (with the longest reaction time and the poorest accuracy). Subjects performed better on the numerical quantifiers with low ranks ("fewer than five," "more than four") than on other determiners, and finally there were no differences between parity quantifiers ("an even number of," "an odd number of") and numerical quantifiers of high rank ("fewer than eight," "more than seven"). In the six-element load condition both numerical and parity quantifiers remained at the same level of difficulty (reaction time and accuracy). However, the performance on proportional quantifiers significantly improved when compared with the four-element working memory load condition. As we will explain this has to do with a trade-off effect between storing and processing.

1.3 The present studies

The aim of our research was to verify the contribution of working memory for specific natural language quantifiers. Data obtained so far by Szymanik and Zajenkowski (2010a, 2010b) support the hypothesis that the complexity of the minimal automata is a good predictor of the difficulty involved in the mental processing of the corresponding quantifier. The present paper extends previous results by studying in greater detail the involvement of working memory in parity and proportional judgments. The results throw some light on the controversy, described in the Introduction, whether from a cognitive perspective, parity quantifiers are more similar to other quantifiers recognizable by finite automata (including elementary quantifiers) or rather to non-elementary proportional quantifiers that correspond to push-down machines.

In the previous investigations of Szymanik and Zajenkowski (2010a, 2010b), proportional quantifiers were the most difficult, while parity quantifiers ranked between proportional and other quantifier types: numerical of small rank or Aristotelian (every, some). Moreover, Szymanik and Zajenkowski (2010b) noticed that under various memory load conditions the relationship between proportional and parity quantifiers varied: under the four-digit memory load they differed significantly in the difficulty, but under a six-digit condition the difference disappeared.

In this paper we report data from two new experiments. In the first study we created a test similar to the span task in Daneman and Carpenter (1980) (cf. Szymanik & Zajenkowski 2010b). We predicted that when subjects are asked to maintain arbitrary information in short-term memory, similar results should be revealed as those described by Szymanik and Zajenkowski (2010a). In particular, the difficulty (indicated by reaction time and accuracy) involved in processing proportional quantifiers should be higher than the difficulty involved in processing parity quantifiers. Additionally, processing of the proportional quantifiers should be influenced to a greater extent by the memory load. The effect should change with respect to the number of elements stored in memory (see Szymanik

& Zajenkowski 2010b). However, the difference between the two classes should be observed under both working memory load conditions and not only in the four-digit memory load situation (cf. Szymanik & Zajenkowski 2010b).

The computational theory of quantifiers (Szymanik 2007) as well as empirical findings (Szymanik & Zajenkowski 2010b; Study 1 in this paper) suggest that proportional quantifiers marshal working memory more than other types of quantifiers. In the second study, we wanted to examine what specific function of working memory is responsible for the difference in processing proportional vs. parity quantifiers. As mentioned above, in working memory one can generally distinguish processing, storage, and a group of executive functions (e.g., supervision and control). The latter two (or one of them) probably determine the uniqueness of proportional quantifier verification: the storage function, because this type of quantifier needs an automaton with a stack, and the executive mechanism responsible for comparing the sizes of two sets. We decided to study the relationship between an individual's storage capacity and the processing of parity and proportional judgments. In order to do this, we presented subjects with two independent tasks: a short-term memory retention task and the quantifier verification task. Next, we investigated the correlation between the memory and verification tasks for proportional and parity quantifiers. If the storage function is more important for the verification of proportional than parity quantifiers we would expect a greater correlation in the case of proportional than parity quantifiers.

2. Study 1

2.1 Participants

Eighty-five native Polish-speaking volunteers from the undergraduate population of two Warsaw universities participated in the study. Of these, 40 were male and 45 were female. The mean age was 22.45 years (SD = 4.5) with a range of 19–40 years. Each subject was tested individually.

2.2 Materials and procedure

In the experiment we used a combined task consisting of two elements. Participants were asked to verify sentences and to memorize a sequence of digits for later recall. The general aim of this study was to assess how subjects under an additional memory load judge the truth-value of statements containing parity and proportional quantifiers.

2.2.1 *Sentence verification*

The task consisted of thirty-two grammatically simple propositions in Polish containing a quantifier that probed a color feature of a car on a display, e.g.:

Więcej	niż połowa	samochodów	jest czerwona.
More than	half of	the cars	are red

Nieparzysta	liczba	samochodów	jest niebieska.
An odd	number of	the cars	are blue.

The same number of color pictures presenting a car park with 15 cars were constructed to accompany the propositions. The colors used for the cars were red, blue, green, yellow, purple and black. Each picture contained objects in two colors (see Figure 3).

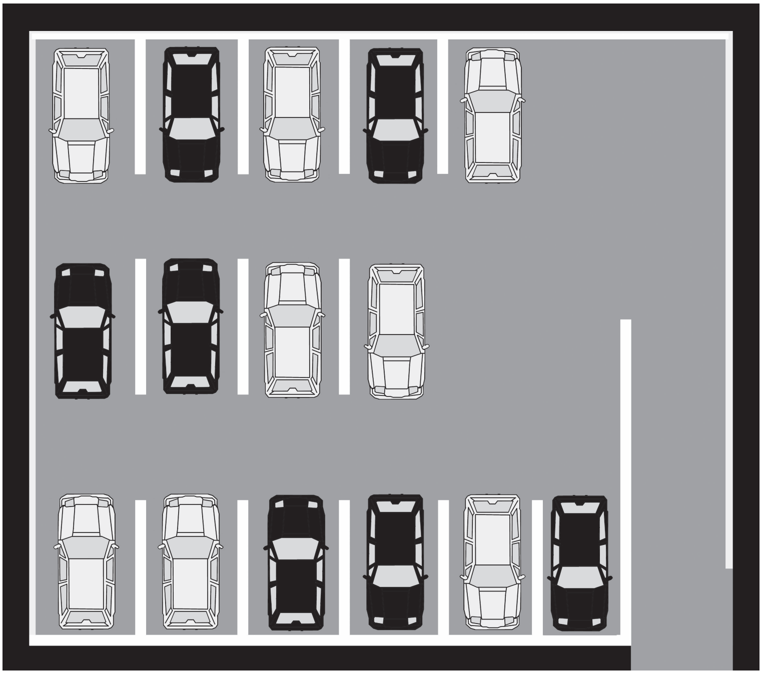

Figure 3. An example of a stimulus used in the first study.

In the study, we used four different quantifiers divided into the following groups:

– parity quantifiers (odd, even), DQs;[2]
– proportional quantifiers (less than half, more than half), PQs.

The groups differ with respect to their complexity: as explained above, parity quantifiers can be recognized by finite automata, while proportional quantifiers necessitate machines with access to unbounded storage. Each quantifier was presented in eight trials. Hence, there were in total 32 trials in the study. Half of each type of items were true and half false. The propositions were accompanied with a quantity of target items near the criterion for validating or falsifying. Therefore, these tasks required a precise judgment (e.g., seven targets in "less than half") (cf. Pietroski, Lidz, Hunter & Halberda 2009). Debriefing following the experiment revealed that none of the participants had been aware that each picture consisted of exactly fifteen objects.

Each quantifier problem involved one 15.5 s event. In an event, a proposition and a stimulus array containing 15 randomly distributed cars were presented for 15,000 ms followed by a blank screen for 500 ms. During the presentation subjects were asked to decide if the proposition accurately described the picture. They responded by pressing the button with the letter "p" if true; the button with the letter "f" was pressed if false. The letters refer to the first letters of the Polish words for "true" and "false."

2.2.2 *Digit recall*

At the beginning of each trial the subjects were presented a sequence of numbers consisting of four or six elements from the range of digits between 0 and 9 and were asked to memorize them. After completing the sentence verification task they were asked to recall the string. Each quantifier type was accompanied by the same number of digits. As correct answers we counted only completely recalled sequences (the right digits in the right order).

2.3 Results

2.3.1 *Sentence verification*

Analysis of variance with repeated measures was used to examine the differences in means in reaction time and the accuracy of the sentence verification task. Each participant had to verify sentences containing one of two types of quantifiers (parity or proportional) while holding in memory strings of digits differing in lengths (four or six elements). Hence, in the study we had the type of quantifier (two levels) and the number of digits (two levels) as two within-subject factors. Each time, all answers from each subject were included in the analysis.

First, we analyzed average reaction times (see Table 1 for means and standard deviations).

Table 1. Mean reaction time (M) and standard deviations (SD) in milliseconds for each quantifier type. "DQs" stands for "parity quantifiers" and "PQs" for "proportional quantifiers."

Quantifier	four-digit		six-digit	
	M	(SD)	M	(SD)
DQs	6768	(1517.44)	6645	(1621.68)
PQs	7634	(1774.27)	7042	(1864.81)

Analysis indicated that the main effects of quantifier type ($F(1, 84) = 14.582$; $p < 0.001$; $\eta^2 = 0.15$) and of number of digits ($F(1, 84) = 9.350$; $p < 0.05$; $\eta^2 = 0.1$) as well as quantifier × digits interaction ($F(1, 84) = 4{,}014$; $p < 0.05$; $\eta^2 = 0.046$) – were significant (see Figure 4).

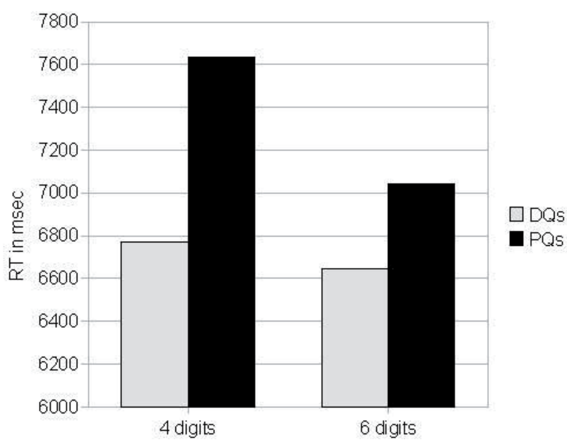

Figure 4. Mean reaction time in four- and six-digit memory load conditions.

For simple effects we analyzed the differences between quantifiers separately for two memory conditions as well as the differences between conditions within the same quantifier type. Pairwise comparison revealed that proportional quantifiers had longer reaction times than parity quantifiers in both memory conditions ($p < 0.05$). Moreover, the average of proportional quantifiers decreased significantly in the six-digit condition but the mean of parity quantifiers remained at the same level between the two conditions.

The total number of correctly verified sentences within each quantifier type and the number of memorized elements was a measure of accuracy in this part of the task. The overall score for each condition ranged from 0 to 8. Table 2 presents the means and standard deviations for quantifier accuracy.

Table 2. Means and standard deviations of the accuracy for each quantifier type.

Quantifier	**four-digit**		**six-digit**	
	M	**(SD)**	**M**	**(SD)**
DQs	6.45	(1.55)	6.48	(1.51)
PQs	5.46	(1.55)	5.94	(1.51)

We found the main effects of quantifier type ($F(1, 84) = 35.296$; $p < 0.001$; $\eta^2 = 0.3$), digits ($F(1, 84) = 4.184$; $p < 0.05$; $\eta^2 = 0.05$) and quantifier × digits interaction (see Figure 5).

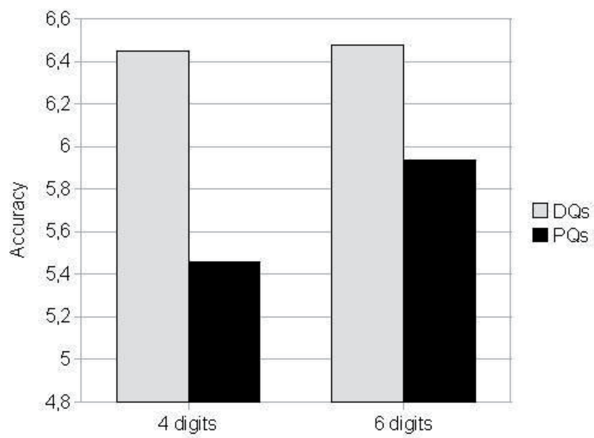

Figure 5. Means and standard deviations of accuracy in four- and six-digit memory load conditions.

In the analysis of simple effects we found that both types of quantifiers differed significantly from each other in the two memory conditions. Between conditions only the performance on proportional quantifiers had changed: the mean was higher in the six-digit situations ($p < 0.05$).

To sum up, the analyses indicated that in both memory conditions proportional quantifiers took longer to solve and accuracy was lower than for parity quantifiers. Additionally, parity quantifiers had equal mean reaction time and accuracy independent of whether the subjects were holding four or six elements in their memory. However, when they were supposed to store six elements in their working memory, subjects performed quantifier tasks with proportional determiners faster and better than in the four-digit memory load situations.

2.3.2 *Digit recall*

ANOVA with two within-subject factors was used to examine how strings of digits (two levels) were recalled with respect to the quantifier type (two levels) with which they were accompanied (see Table 3 for means and standard deviations).

Table 3. Means and standard deviations of the accuracy for each type of quantifier.

Quantifier	four-digit		six-digit	
	M	(SD)	M	(SD)
DQs	6.22	(1.85)	3.94	(2.66)
PQs	5.32	(1.92)	4.01	(2.43)

The analysis indicated the main effect of digits ($F(1, 84) = 8.863$; $p < 0.004$; $\eta^2 = 0.095$), quantifiers ($F(1, 84) = 73.533$; $p < 0.001$; $\eta^2 = 0.47$) and digits × quantifier interaction ($F(1, 84) = 13.481$; $p < 0.001$; $\eta^2 = 0.14$) (see Figure 6).

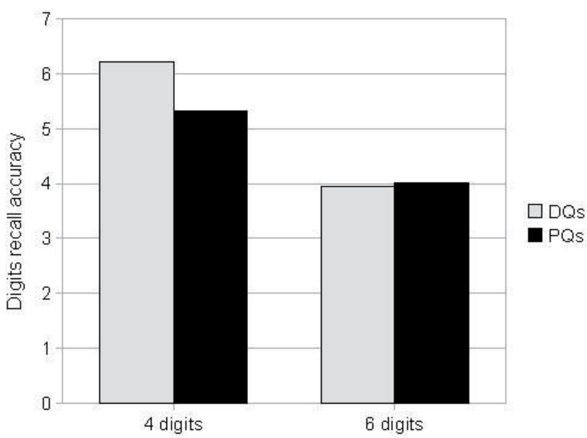

Figure 6. Accuracy of four- and six-digit recall with respect to the quantifier types.

To examine the interaction effect we compared recall accuracy for four and six digits as well as performance on digit recall with respect to quantifier types separately for the two conditions. We found that six elements were memorized worse than four digits. Moreover, four digits accompanying proportional quantifiers were recalled more poorly than with parity determiners, but in the six-element condition there were no differences.

3. Study 2

In the second study we tried to examine the role of short-term retention of information, a component of working memory, in parity and proportional quantifiers. To capture an individual's pure storage capacity we presented the short-term memory (STM) task independently from the sentence verification test.

3.1 Participants

Thirty-eight native Polish-speaking volunteers from the undergraduate population of two Warsaw universities participated in the study. Of these, 18 were male and 20 were female. The mean age was 24.0 years (SD = 4.68) with a range of 18–30 years. Each subject was tested individually.

3.2 Materials and procedure

The subjects performed two independent tasks. The first one was identical with the sentence verification part described in Study 1. The second task was a computerized version of Sternberg's short-term memory measure (Sternberg 1966). On each trial of the test, the subjects were presented with a random series of different digits, one at a time, for 300 ms, followed by a blank screen and the test digit. Participants had to decide whether the test digit had appeared in the previously displayed string. Sequences of digits of three lengths (four, six, or eight) were repeated eight times each; hence, there were 24 trials overall.

3.3 Results

Pearson's correlations between each condition of memory task and quantifier accuracy and RT were computed (see Table 4). The only significant results were related to the eight recalled items and the accuracy of quantifiers. Parity and proportional quantifiers were positively associated with the memory task, which means that the higher the memory score, the better the performance with both quantifiers. The magnitudes of the two significant correlations were very similar, showing R^2 equal to 14.5% for proportional and 11.6% for parity quantifiers. This means that the storage function explains almost the same percentage of the variance in both quantifiers.

Table 4. Pearson's correlations between the memory task and quantifier verification task ($N = 38$).

Memory task	Quantifier accuracy		Quantifier RT	
	DQs	PQs	DQs	PQs
four items	0.18	0.21	0.24	0.24
six items	0.12	0.77	–0.25	–0.05
eight items	0.34*	0.38*	–0.16	0.03

* p values < 0.05 (two-tailed)

4. Discussion

In the first study we examined the role of working memory in parity and proportional quantifier verification. Our results show that under additional memory load, proportional quantifiers take longer to process, and with poorer results, than parity quantifiers. This is consistent with the earlier research of Szymanik and Zajenkowski (2010a, 2010b). In both memory conditions, parity quantifiers were processed equally long and with the same accuracy, while in the case of proportional quantifiers, in the six-digit situation the accuracy increased and RT decreased. Moreover, four digits were recalled worse when they accompanied proportional sentences than parity sentences, but there were no differences between quantifiers while subjects were storing six elements.

Taking this data all together, we begin by assuming that holding six digits in memory while simultaneously solving the verification task was too difficult for subjects regardless of quantifier type, and probably they gave up in the memory task. That is why we obtained equally low means in the six-digit situation. In that condition, when people didn't really focus on the memory task and had a "free stack," they performed better on proportional quantifiers (they were no longer distracted), but not on parity quantifiers (which were processed similarly either with full or free memory storage). Hence, the instruction to hold either four or six digits in memory makes no difference to the processing of parity quantifiers, but has an impact on proportional quantifiers.

That is why our experiment sheds some more light on the nature of proportional quantifiers. As we mentioned above this type of quantifier demands a recognition mechanism with unbounded internal memory. It involves tracking and comparing the relative sizes of two sets. Let us consider an example. To verify a sentence "More than half of the cars are red," one has to count and hold in short-term memory the number of red cars and then compare it with the number of all cars (or with the number of blue cars). No such memory storage is necessary in the processing of parity quantifiers; here a subject needs only to focus on counting elements of just one set, e.g., red cars, and there is no need for an extensive memorization/comparison process to be involved.

Moreover, the second study showed that a larger storage capacity facilitates the processing of proportional and parity quantifiers to the same extent. The relationships between the accuracy of the two analyzed quantifier types and short-term memory were very similar. In other words, the efficiency of the verification of the two quantifiers depends on the storage function in the same way. Therefore, we have reasons to think that, from the psychological viewpoint, it is not the storage function but rather some managing and executive functions that are responsible for the uniqueness of proportional sentence verification.

The observations made in the previous paragraphs are crucial for understanding our results. We believe that the differences obtained between parity and proportional quantifiers are due to the different cognitive functions involved in their verification. Parity quantifiers require only the processing function, understood as the transformation of information,[3] but proportional quantifiers also necessitate temporary storage and some executive processes, such as comparison. Hence, keeping arbitrary data in mind interferes with the process of verifying proportional quantifiers and impairs their comprehension to a higher degree. Let us give two explanations of this phenomenon.

In line with the computational theory of quantifier verification we would say that elements from the memory task are held in the storage unit (stack) where the sentence verification also needs to place some information. The storage system has a limited capacity and, therefore, not all elements can be stored and subsequently processed. This mechanism hinders the parallel solving of the memory task and the proportional sentence verification task. However, in the case of parity quantifiers, there is no conflict.

The second explanation refers to supervision as an executive function of working memory. In the theory of Oberauer et al. (2003), described previously, supervision is defined as the selective activation of relevant representations and the prevention of distraction. This definition is based on earlier work, Miyake et al. (2000), who found a common factor in such functions as shifting between tasks, inhibiting dominant responses, or updating and controlling the ongoing cognitive processes. It is possible that supervision is a crucial function when it comes to verification of the proportional quantifiers. In that process one has to keep track of the size of one set while counting elements from the second set, which could be distracting. To succeed, the cognitive system has to recruit the supervision function. Presumably, it was especially important in our first study, in which participants had to hold in memory additional information plus additional distractors, while solving the quantifier task.

Future research focusing on the psychological modeling of quantifier comprehension could help clarify the relationship between those two, not necessarily mutually exclusive, explanations. First of all, we should identify cognitive correlates of different aspects of computational models, such as the difference between states (counting) and stack (counting + storage + comparison). The aim would be to pin down the specific cognitive mechanisms responsible for quantifier comprehension, taking into account factors such as the role of supervision, attentional costs, and storage functions. Finally, we plan to design an experiment differentiating between storage and supervision costs in quantifier processing. For instance, a version of the dual task setup could be used, where memorizing digits is replaced by a task engaging supervision rather than storage, such as counting dots in a visual array. We believe that research like this may not only help to better understand the psychological plausibility of semantic theories but also enrich our knowledge of specific cognitive mechanisms constraining working memory.

Acknowledgments

The first author was supported by a Vici grant NWO-277-80-001. The second author was supported by a program run by the Foundation for Polish Science and grant no. IP2010 048670 of the Polish Ministry of Science and Higher Education.

Notes

1. See also the discussion between Szymanik and Zajenkowski (2009) and Troiani, Peelle, Clark and Grossman (2009).

2. "DQ" stands for divisibility quantifiers.

3. Recall the definitions of working memory functions from Section 1.1.

References

Baddeley, A. 1986. *Working Memory*. Oxford: Oxford University Press.

Baddeley, A. & Hitch, G. 1974. "Working memory". In G. Bower (ed.), *The Psychology of Learning and Motivation*. New York: Academic Press, 47–90.

Benthem, J. van. 1986. *Essays in Logical Semantics*. Dordrecht: Reidel.
Clark, H. & Chase, W. 1972. "On the process of comparing sentences against pictures". *Cognitive Psychology* 3(3): 472–517.
Conway, A. & Engle, R. 1996. "Individual differences in working memory capacity – more evidence for a general capacity theory". *Memory* 6: 122–125.
Cowan, N. 1995. *Attention and Memory: An Integrated Framework*. Oxford: Oxford University Press.
Daneman, M. & Carpenter, P. 1980. "Individual differences in working memory and reading". *Journal of Verbal Learning and Verbal Behavior* 19: 450–466.
Daneman, M. & Green, I. 1986. "Individual differences in comprehending and producing words in context". *Journal of Memory and Language* 25: 1–18.
Duff, S. & Logie, R. 2001. "Processing and storage in working memory span". *The Quarterly Journal of Experimental Psychology* 54: 31–48.
Engle, R., Kane, M. & Tuholski, S. 1999. "Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex". In A. Miyake & P. Shah (eds), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. London: Cambridge University Press, 102–134.
Hopcroft, J. E., Motwani, R. & Ullman, J. D. 2000. *Introduction to Automata Theory, Languages, and Computation*. Reading, MA: Addison Wesley, 2nd edition.
Just, M. & Carpenter, P. 1992. "A capacity theory of comprehension: Individual differences in working memory". *Psychological Review* 99: 122–149.
King, J. & Just, M. 1991. "Individual differences in syntactic processing: The role of working memory". *Journal of Memory and Language* 30: 580–602.
McMillan, C., Clark, R., Moore, P., Devita, C. & Grossman, M. 2005. "Neural basis for generalized quantifier comprehension". *Neuropsychologia* 43: 1729–1737.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A. & Wager, T. D. 2000. "The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis". *Cognitive Psychology* 41: 49–100.
Moxey, L. M. & Sanford, A. J. 1993. *Communicating Quantities: A Psychological Perspective*. Hove: Lawrence Erlbaum Associates.
Oberauer, K., Suess, H., Wilhelm, O. & Wittman, W. 2003. "The multiple faces of working memory: Storage, processing, supervision, and coordination". *Intelligence* 31: 167–193.
Oberauer, K., Suess, H., Wilhelm, O. & Wittman, W. 2008. "Which working memory functions predict intelligence?". *Intelligence* 36: 641–652.
Peters, S. & Westerståhl, D. 2006. *Quantifiers in Language and Logic*. Oxford: Oxford University Press.
Pietroski, P., Lidz, J., Hunter, T. & Halberda, J. 2009. "The meaning of 'most': Semantics, numerosity, and psychology". *Mind and Language* 24(5): 554–585.
Sternberg, S. 1966. "High-speed scanning in human memory". *Science* 153: 652–654.
Suess, H., Oberauer, K., Wittman, W., Wilhelm, O. & Schulze, R. 2002. "Working-memory capacity explains reasoning ability – and a little bit more". *Intelligence* 30: 261–288.
Szymanik, J. 2007. "A comment on a neuroimaging study of natural language quantifier comprehension". *Neuropsychologia* 45: 2158–2160.
Szymanik, J. 2009. *Quantifiers in TIME and SPACE. Computational Complexity of Generalized Quantifiers in Natural Language*. PhD thesis, Universiteit van Amsterdam.

Szymanik, J. & Zajenkowski, M. 2009. "Improving methodology of quantifier comprehension experiments". *Neuropsychologia* 47: 2682–2683.
Szymanik, J. & Zajenkowski, M. 2010a. "Comprehension of simple quantifiers. Empirical evaluation of a computational model". *Cognitive Science* 34: 521–532.
Szymanik, J. & Zajenkowski, M. 2010b. "Quantifiers and Working Memory". In M. Aloni & K. Schulz (eds), *Amsterdam Colloquium 2009*. Berlin: Springer-Verlag, 456–464. (*Lecture Notes in Artificial Intelligence*, 6042).
Troiani, V., Peelle, J., Clark, R. & Grossman, M. 2009. "Magnitude and parity as complementary attributes of quantifier statements". *Neuropsychologia* 47: 2684–2685.
Turner, M. L. & Engle, R. 1989. "Is working memory capacity task dependent?". *Journal of Memory and Language* 28: 127–154.
Unsworth, N., Heitz, R., Schrock, J. & Engle, R. 2005. "An automated version of the operation span task". *Behavior Research Methods* 37: 498–505.