Monotonicity has only a relative effect on the complexity of quantifier verification

Szymanik, J.; Zajenkowski, M.

Publication date
2013

Document Version
Final published version

Published in
Proceedings of the 19th Amsterdam Colloquium

Monotonicity has only a relative effect on the complexity of quantifier verification

Jakub Szymanik* and Marcin Zajenkowski†

[1] Institute for Logic, Language and Computation
University of Amsterdam
J.K.Szymanik@uva.nl
[2] Faculty of Psychology, University of Warsaw
Zajenkowski@psych.uw.edu.pl

Abstract

We discuss a computational model of quantifier verification. It predicts that there is *no effect of monotonicity* on the verification of numerical quantifiers but *only the interaction of monotonicty and sentential truth-values*. Moreover, it predicts no monotonicity or interaction with truth-values effects for proportional quantifiers. We present an experimental study supporting the predictions of the computational model. We argue that the role of the interaction between monotonicty and sentential truth-values as well as the differences between various quantifier classes (numerical vs. proportional) have been regularly overlooked in the literature.

1 Introduction

Monotonicity is considered to be one of the key properties of languages both in logic and linguistics. Barwise and Cooper [1] even suggested that monotonicity is one of the *semantic universals*: the simplest noun phrases of any natural language express monotone quantifiers or conjunctions of monotone quantifiers. They have also noted that monotonicity relates to the intuitive truth-value checking procedures for quantified sentences. For example, imagine a parking lot filled with cars. To verify an upward monotone (increasing) sentence 'More than seven cars are green', you need to find at least eight green cars (a so-called witness set in Barwise's and Cooper's terminology). For a downward monotone (decreasing) sentence, e.g., 'Fewer than eight cars are red' you must check all the cars and make sure that there are no more than seven green cars. Based on the intuitive complexity of these procedures, Barwise and Cooper predicted that *'response latencies for verification tasks involving decreasing quantifiers would be somewhat greater than for increasing quantifiers' (p. 192)*.

It seems that [1] has overlooked the truth-value of the sentence as an important aspect of the verification complexity. For instance, if the upward monotone sentence is true, then indeed one needs

*Acknowledges the support of Veni grant NWO-639.021.232
†Acknowledges the support of Polish National Science Centre grant no. 2011/01/D/HS6/01920.

to just find any witness set; otherwise, when the sentence is false, one needs to check all the cars and make sure that really not more than 7 of them are green. If you can perceptually quickly identify the set satisfying the predicate in question, e.g., the set of all green cars, then for true upward monotone quantifier it will take more counting than for the corresponding false quantifier to judge whether the set is 'large enough'. However, for the downward monotone sentence the situation is exactly the opposite. Moreover, intuitively, false (resp. true) instances of an upward monotone sentence are equally hard as true (resp. false) instances of an downward monotone sentence. Therefore, it seems that we should rather expect *the effect of interaction between monotonicity and truth-value* than pure monotonicity to have an impact on the verification complexity.

Thinking about quantifier verification in terms of computations [2, 8] can help us to clarify the intuitions about the influence of monotonicity on verification, and its interaction with truth-values:

Sentences with upward monotone numerical quantifiers, e.g., 'more than seven', should take longer to process when they are true than when they are false. For true 'more than seven' subjects need to count up to eight while for the false sentences only up to seven. In case of the sentences' containing downward monotone numerical quantifiers, like 'fewer than eight', the relationship is reversed. This hypothesis directly improves on [1] by taking into account the interaction between monotonicity and truth-value.

Proportional quantifiers should be more difficult, as the minimal verification procedure triggered by them is cognitively more complex [6, 7, 9–13]. It should be the case even in our experimental setting where 'more than half' ('less than half') and 'more than seven' ('less than eight') are denotationally equivalent. The reason is that the corresponding procedure is triggered automatically rather by the linguistic form than the situation to be judged. Moreover, in the case of proportional quantifiers, subjects need to always compare all elements, no matter whether the sentences are true or false, and so there should be no significant difference in the processing difficulty between the upward and the downward monotone proportional quantifiers, not even when taking into account truth-values.

2 Experiment

Sixty nine native Polish-speaking adults took part in the study (31 male and 38 female). They were volunteers from the University of Warsaw undergraduate population. The mean age was 21.42 years (SD = 3.22) with a range of 18–30 years. Each subject was tested individually and was given a small financial reward for participation in the study.

The task used in the study consisted of sixteen grammatically simple sentences in Polish, containing a quantifier that probed a color feature of a car in a display. Each picture contained fifteen objects in two colors. Four different quantifiers were presented to each subject in four trials. The quantifiers were: fewer than eight, more than seven (numerical of high rank), fewer than half, more than half (proportional).

For each quantifier, half of the sentences were true. The sentences were accompanied by pictures with a quantity of target items near the criterion for validating or falsifying the proposition, therefore requiring a precise judgment (e.g., seven targets in 'fewer than half'). In each quantifier problem, first the proposition appeared in the middle of a screen, followed by a blank screen (500 ms) and stimulus array containing 15 randomly distributed cars. Hence, within each trial, the sentence and the picture were presented separately. We recorded then the time used for reading the sentence and the time used for verification with the picture. Debriefing, that followed the experiment, revealed that none of the participants were aware that each picture consisted of fifteen objects.

The stimulus arrays were presented for 15000 ms. Within this time, the subjects were asked to decide if the proposition accurately described the presented picture. They responded by pressing the buttons referring to the first letters of the Polish words for 'true' ('p') and 'false' ('f'). All stimuli were counterbalanced and randomly distributed throughout the experiment.

2.1 Results

In this study, the sentence was presented before the picture was displayed, hence the true-false conditions were not included in the analysis. ANOVA with type of quantifier (2 levels: numerical, proportional), and monotonicity (2 levels: upward, downward) as the two within-subject factors was used to examine the differences in mean reaction times of sentence mean reading time (see Table 1 for means and standard deviations). There were no significant effects or interactions between the factors.

Table 1: Means (M) and standard deviations (SD) of the reading time in milliseconds of each quantifier

Quantifier	M	SD
More than seven	4054	1992
Fewer than eight	4345	2913
More than half	4459	2907
Fewer than half	4742	2863

Further, the times needed to verify the quantifiers were compared. ANOVA with type of quantifier (2 levels: numerical, proportional), monotonicity (2 levels: upward, downward), and the statement's truth-value (2 levels: true, false) as three within-subject factors was used to examine differences in mean reaction times of sentence-picture verification (see Table 2).

The analysis indicated significant main effects of quantifier ($F(1, 68) = 146.73$, $p < 0.001$, pη^2=0.68) and monotonicity ($F(1, 68) = 6.73$, $p = 0.012$, pη^2=0.09), as well as the following interactions: quantifier × monotonicity ($F(1, 68) = 13.32$, $p < 0.001$, pη^2=0.16), quantifier × truth ($F(1, 68) = 11.58$, $p = 0.005$, pη^2=0.15), monotonicity × truth ($F(1, 68) = 7.93$, $p = 0.006$, pη^2=0.10), and quantifier × monotonicity × truth ($F(1, 68) = 12.64$, $p < 0.001$, pη^2=0.16) (see Figure 1).

Analyzing the latter interaction effect, we compared the differences within each type of quantifier. We performed ANOVA with monotonicity and truth-value as two within-subject factors for the numerical and proportional quantifiers separately.

In the case of the numerical quantifiers, we found significant effects of monotonicity ($F(1, 68) = 43.61$, $p < 0.001$, pη^2=0.40), truth ($F(1, 68) = 13.01$, $p = 0.001$, pη^2=0.16) and the interaction monotonicity × truth ($F(1, 68) = 37.54$, $p < 0.001$, pη^2=0.36). Pairwise comparisons among means

Table 2: Means (M) and standard deviations (SD) of the verification time in milliseconds of each quantifier with respect to monotonicity and truth-value.

Quantifier		M	SD
More than seven	True	3793	1241
	False	3360	1218
	Overall	3577	964
Fewer than eight	True	3626	1296
	False	5029	1833
	Overall	4327	1307
More than half	True	6511	2454
	False	6475	2195
	Overall	6493	1959
Fewer than half	True	6509	2378
	False	6084	2464
	Overall	6296	2090

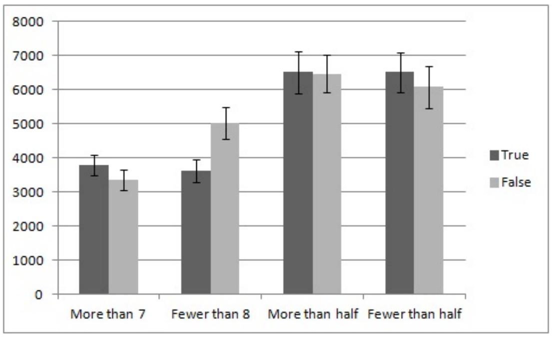

Figure 1: Average reaction time in milliseconds of each experimental condition. *Note* Error bars are for 95% intervals.

(with LSD test) revealed that the false 'fewer than eight' were processed longer than any other quantifier, while false 'more than seven' were performed the fastest. Moreover, there was no significant difference between both true conditions ($p > 0.05$).

There were no significant differences within proportional quantifiers ($p > 0.05$).

We also analyzed the main effect of the quantifier type. The analysis revealed that proportional quantifiers (M=6395, SD=1832) were processed longer than numerical (M=3952, SD=1069).

The accuracy analysis revealed only one significant effect of quantifier type ($F(1,\ 68) = 20.23$, $p < 0.001$, pη^2=0.23): proportional quantifiers were more difficult (M=.85, SD=.02) than numerical quantifiers (M=.94, SD=.01).

Summing up, we can conclude that there are no differences within proportional quantifiers as far as the monotonicity and truth is concerned. With regards to the numerical sentences, their difficulty increased as follows: false ‘more than seven’, both true ‘more than seven’ and ‘fewer than eight’ (equal), and false ‘fewer than eight’. Finally, the proportional statements were generally more difficult than the numerical.

3 Discussion

We have examined the time needed to process sentences containing quantifiers. The quantifiers we have studied differed in their computational complexity, monotonicity, and true-false conditions. The results have confirmed the complexity hypothesis derived from the computational model. As predicted, sentences with the quantifier ‘more than seven’ were processed faster when they were false. In the case of ‘fewer than eight’ true sentences were easier to process than false sentences. In other words, we found an interaction effect between monotonicity and truth-value that reflects the effect of counting in the case of numerical quantifiers. Moreover, our data indicated that there was no significant monotonicity effect within the proportional quantifiers. Again, this is in agreement with the computational theory, according to which the mental strategies for the verification of proportional quantifiers resemble the push-down automata algorithm [2].

The running of the procedure does not differ between the upward monotone case ‘more than half’ and the corresponding downward monotone quantifier ‘less than half’. Furthermore, the complexity of the computation is similar between true and false instances of the proportional sentences. In both cases, one needs to compute and compare the cardinalities of two sets that cover the whole universe. These two facts explain why we found no effect of monotonicty or truth-value in the case of proportional quantifiers.

The average difference in reaction time was additionally consistent with the hypothesis that quantifiers trigger the corresponding minimal computation (counting up to seven or eight). Therefore, our research contributes another argument in favor of the cognitive plausibility of the automata-theoretic model of quantifier verification [2]. Furthermore, we observed that, in general, proportional quantifiers are more difficult than numerical quantifiers, which is again consistent with the theory that predicts a complexity difference between numerical and proportional quantifiers.

As we have already extensively justified throughout the introduction, our computational perspective brings a refinement to the theory proposed in [1]. Namely, we predicted and experimentally found an interaction effect between monotonicity and truth value in the case of numerical quantifiers. This effect follows from the corresponding differences in their computational complexity. Moreover, neither our theory nor the experiments indicate an involvement of monotonicity in the difficulty of proportional quantifier verification. These observations allow suggesting that monotonicity has only a relative effect on the difficulty of verification. Together with other aspects of the situation, monotonicty may influence the complexity of the verification.

In a syllogistic reasoning experiment, [4] has found that if the monotonicity profiles of two quantifying expressions are the same, then they should be equally hard to process. They studied sentences like:

(1) Some of the sopranos sang with more than three of the tenors.

(2) None of the sopranos sang with fewer than three of the tenors.

(3) Some of the sopranos sang with fewer than three of the tenors.

The results suggested that sentences with two upward monotone quantifiers, like (1), are easier to reason with than sentences with two downward monotone quantifiers, e.g., (2). While sentences with two quantifiers of different monotonicity, for instance sentence (3), are the hardest. According to our experiments, in the verification tasks, quantifiers with the same monotonicity profiles can differ with respect to their difficulty depending on the truth-value. This suggest that there might be crucial differences between reasoning with quantifiers and quantifier verification and that future studies devoted to monotonicity should control the true-false conditions and include them in the analyses.

Moreover, our results complement the proposal put forward in [5] that the counting stage in processing is affected by the number mentioned in the quantifier, rather than the critical number of objects needed to verify the statement. They reported a real time study of verification procedures for numerical quantifiers, like 'more than n' and 'fewer than n', using self-paced counting. The used methodology is an analogue of well-known self-paced reading experiments. Subjects hear a sentence and are asked to determine as fast and as reliably as possible its truth-value relative to an array of dots. The arrays are presented as three scattered rows of hexagonal plates. As participants press the space bar, the dots are uncovered in groups of 1, 2, or 3, while previously seen dots are recovered and masked. Participants may answer once they have enough information. The setting allows looking into the verification process by timing how the participants uncover the dots. Using this paradigm, [5] was able to show that reaching the number heard in the quantifier causes a slow down in the processing. This observation is consistent with our data emphasizing the interaction with the truth-value, as changes in the truth-value are necessarily bound with reaching the number n. This also suggests that next to the reading and verification stages, one should also take into account the *decision stage* in quantifier verification.[1] One would predict that the interactions among all three processing stages: reading, verification, and decision, may, for instance, play a crucial role in explaining the differences between comparative quantifiers, e.g., 'more than 3' and the equivalent superlative quantifiers, e.g., 'at least 4'. We know that superlative quantifiers are harder to verify than the corresponding comparative quantifiers but there are no differences in reading times [3]. Our computational approach predicts no differences in the verification times as the counting processes for equivalent quantifiers are identical, i.e., the computations for 'more than 3' and 'at least 4' do not differ in complexity. Therefore, we would predict that the difference between comparative and superlative quantifiers is due to the decision stage. This prediction falls outside the scope of the current work but should be tested in the future experiments.

[1]Interestingly, [5] chose the quantifiers in a way that no matter whether the sentence was true or false, the subjects always needed to count only up to seven, i.e., the number heard, n, varied across true and false items, e.g., 'more than six' (true) but 'more than seven' (false) and 'fewer than eight' (true) but 'fewer than seven' (true). As a result they found out that monotone increasing quantifiers are quicker to verify than falsify and monotone decreasing quantifiers are quicker to falsify than verify.

References

[1] Jon Barwise and Robin Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4:159–219, 1981.

[2] Johan van Benthem. *Essays in Logical Semantics.* D. Reidel, Dordrecht, 1986.

[3] Bart Geurts, Napoleon Katsos, Chris Cummins, Jonas Moons, and Leo Noordman. Scalar quantifiers: Logic, acquisition, and processing. *Language and Cognitive Processes*, 25(1):244–253, 2009.

[4] Bart Geurts and Frans van der Silk. Monotonicity and processing load. *Journal of Semantics*, 22:97–117, 2005.

[5] Jorie Koster-Moeller, Jason Varvoutis, and Martin Hackl. Verification procedures for modified numeral quantifiers. In N. Abner and J. Bishop, editors, *Proceedings of the 27th West Coast Conference on Formal Linguistics*, volume 1986, pages 310–317, Somerville, 2008. Cascadilla Proceedings Project.

[6] Corey T. McMillan, Robin Clark, Peachie Moore, Christian Devita, and Murray Grossman. Neural basis for generalized quantifier comprehension. *Neuropsychologia*, 43:1729–1737, 2005.

[7] Corey T. McMillan, Robin Clark, Peachie Moore, and Murray Grossman. Quantifiers comprehension in corticobasal degeneration. *Brain and Cognition*, 65:250–260, 2006.

[8] Jakub Szymanik. *Quantifiers in TIME and SPACE. Computational Complexity of Generalized Quantifiers in Natural Language.* PhD thesis, University of Amsterdam, Amsterdam, 2009.

[9] Jakub Szymanik and Marcin Zajenkowski. Comprehension of simple quantifiers. Empirical evaluation of a computational model. *Cognitive Science: A Multidisciplinary Journal*, 34(3):521–532, 2010.

[10] Jakub Szymanik and Marcin Zajenkowski. Quantifiers and working memory. In M. Aloni and K. Schulz, editors, *Amsterdam Colloquium 2009, Lecture Notes In Artificial Intelligence 6042*, pages 456–464. Springer, 2010.

[11] Jakub Szymanik and Marcin Zajenkowski. Contribution of working memory in parity and proportional judgments. *Belgian Journal of Linguistics*, 25(1):176–194, 2011.

[12] Marcin Zajenkowski, Rafał Styła, and Jakub Szymanik. A computational approach to quantifiers as an explanation for some language impairments in schizophrenia. *Journal of Communication Disorders*, 44(6):595 – 600, 2011.

[13] Marcin Zajenkowski and Jakub Szymanik. Most intelligent people are accurate and some fast people are intelligent: Intelligence, working memory, and semantic processing of quantifiers from a computational perspective. *Intelligence*, 41(5):456 – 466, 2013.