Alternative Representations in Formal Semantics: A case study of quantifiers

Shane Steinert-Threlkeld[1], Gert-Jan Munneke[2], and Jakub Szymanik[2]

[1] Department of Philosophy, Stanford University
shanest@stanford.edu
[2] Institute for Logic, Language, and Computation, University of Amsterdam
G.J.Munneke@uva.nl, J.K.Szymanik@uva.nl

Abstract

How do semantic theories fit into the psychology of language more generally? A number of recent theoretical and experimental findings suggest that specifications of truth-conditions generate biases for different verification procedures. In this paper, we show how considerations of different representations of a visual scene in the semantic automata framework can generate predictions for differential working memory activation in proportional quantifier sentence verification. We present experimental results showing that different representations do impact working memory in sentence verification and that 'more than half' and 'most' behave differently in this regard.

1 Introduction

Semantic theories for (fragments of) natural languages typically provide a compositional assignment of truth-conditions to sentences in the repsective language. Such assignments are used to predict competent speakers' judgments of entailment and the truth of sentences in context. This level of generality, however, leaves open exactly how a specification of truth-conditions integrates with cognition more broadly to generate these judgments. There are at least two natural views to take on this issue. The relationship may be *permissive*: once a specification of truth-conditions is "exported" to general cognition, anything goes. There is no systematic connection between the ways that truth-conditions are specified and judgments of truth in context are made. On the other hand, the relationship may be *constrained*: the ways in which truth-conditions are specified correlates with and constrains the methods of verification of sentences in context.

A number of philosophers of language, developing ideas rooted in Frege, have argued that knowing the meaning of a sentence consists in having 'internalized' an algorithm for computing the truth-value of that sentence in a context [Dummett, 1973, Suppes, 1980, Moschovakis, 2006, Horty, 2007]. This line of thought has been put to experimental test recently by psychologists and linguists [Hackl, 2009, Pietroski et al., 2009, Lidz et al., 2011]. These experiments have led the theorists to argue that the relationship between specifications of truth-conditions and verification procedures is in fact constrained. The evidence comes in two forms. In one case, subjects are asked to verify the truth of a single sentence against visual scenes which differ in how amenable they are to different verification procedures. [Pietroski et al., 2009] and [Lidz et al., 2011] find that in the case of sentences involving 'most', such manipulations do not effect verification accuracy. This suggests that the specification of truth-conditions for 'most' constrains in some way the verification procedures available.[1] In another case, subjects

[1]In particular, [Lidz et al., 2011] propose the *Interface Transparency Thesis*: "speakers exhibit a bias towards the verification procedures provided by canonical specifications of truth conditions" (p. 229). While we find this thesis imminently plausible, we keep the discussion at a more general level for reasons that will become clear in the later discussion of results.

are asked to verify two truth-conditionally equivalent sentences against visual scenes in the same experimental paradigm. [Hackl, 2009] finds that there are differences in the way subjects perform self-paced counting tasks when verifying sentences containing 'most' and 'more than half'. Since the sentences are truth-conditionally equivalent, this suggests that the two quantifiers possess different specifications of the truth-conditions, which constrain the methods of verification.[2]

In this paper, we contribute to the growing body of evidence for a constrained relationship by exploring the impact of different presentations of a visual scene on working memory load in proportional quantifier sentence verification. First, we present a computational model for quantifier meanings which has made empirically verified predictions about working memory in sentence verification. Then, we show how to extend that framework to handle different representations of the visual scene. This extension motivates a prediction that how a visual scene is presented will effect working memory involvement. We then present experimental results that make good on this prediction. The results also indicate that 'most' and 'more than half' are effected differently by the visual scene manipulation, providing further evidence for a constrained relationship between truth-condition specifications and verification procedures. The paper concludes with a discussion of future directions.

2 Semantic Automata and Working Memory

The semantic automata approach to generalized quantifiers associates with each quantifier Q a formal language L_Q and a machine M_Q accepting L_Q. Definability of quantifiers in various logics corresponds to levels of the Chomsky hierarchy of languages/machines. All first-order definable quantifiers have a finite-state automaton (FSA) / regular language, while others – notably proportional quantifiers like *most* and *more than half* – require a pushdown automaton (PDA) / context-free language [van Benthem, 1986, Mostowski, 1998]. A pushdown automaton essentially augments a finite-state automaton with a stack, a form of *memory.* If one views the automata as verification procedures somehow internalized in the minds of competent speakers of the language, this leads one to expect that verifying sentences with quantifiers which require a PDA will use more working memory than verifying sentences with quantifiers that have FSAs.

This hypothesis has been demonstrated true in a number of experiments over the past decade. In a pioneering study, [McMillan et al., 2005] had participants verify sentences of the form 'Q of the balls are blue' while in an fMRI machine. They varied Q between quantifiers that have FSAs ('some', 'all', 'at least three') and those that only have PDAs ('most'). In addition to behaviorally finding that the latter are more difficult to verify, they found differential activation in dorsolateral prefrontal and inferior frontal cortices bilaterally. These brain regions have previously been found to be highly involved in working memory by, among others, [Braver et al., 1997].[3]

[2] An apparently related line of work concerns the relationship between superlative quantifiers, such as 'at least n', and the corresponding comparative quantifier 'more than $n-1$'. Traditional semantic theories, such as generalized quantifiers theor, consider these to be truth-conditionally equivalent. [Geurts et al., 2010] show experimentally that the two quantifiers are processed differently. They, however, build on [Geurts and Nouwen, 2007], in which it is argued that the two quantifiers do in fact differ in subtle ways in their truth-conditions. Nevertheless, [Cummins and Katsos, 2010] provide further experiments in support of the view that superlatives and comparatives are truth-conditionally equivalent but have different psychological profiles due to the presence of strict versus weak inequality in the specification of the truth-conditions. Under this interpretation, these quantifiers exhibit a similar phenomenon to the one being discussed here.

[3] This line of work has continued in [McMillan et al., 2006, Troiani et al., 2009a, Troiani et al., 2009b] and has been nicely summarized by [Clark and Grossman, 2007, Clark, 2011].

This differential activation of working memory in quantifier verification has also been demonstrated behaviorally. [Szymanik and Zajenkowski, 2010] show that reaction times in a sentence verification task are much higher for proportional quantifiers than for first-order ones. [Szymanik and Zajenkowski, 2011] combined quantifier sentence verification with a memory span task. In particular, participants were asked to memorize a sequence of either 4 or 6 digits before performing the verification task and were then asked to recall it afterwards. They found that reactions times were much longer and accuracy much worse in verification for proportional quantifiers than for FSA quantifiers.[4] Moreover, in the 4-digit case, digit recall performance was significantly lower after proportional quantifier verification. Finally, [Zajenkowski et al., 2011] have compared the verification of natural language quantifier sentences in a group of patients with schizophrenia (and associated working memory deficits) and a healthy control group. In both groups, the difficulty of the quantifiers was consistent with the computational predictions. Patients with schizophrenia took more time to solve the problems with every quantifier. They were significantly less accurate, however, only with proportional quantifiers.[5] These results, together with the neuroimaging ones, show that the semantic automata model makes good predictions about working memory involvement in quantified sentence verification.

2.1 Representation in Semantic Automata

All of the aforementioned work, however, neglects a crucial component of the psychological task: *how* the visual scene is represented. To show how the semantic automata model can also be used to make predictions about the effect of visual presentation, we must explain the model in slightly more detail. A generalized quantifier Q is a class of finite models of the form $\langle M, A, B\rangle$ [Barwise and Cooper, 1981]. To generate L_Q, a mapping $\tau(\cdot)$ from such models into strings of 0s and 1s is defined by sending elements of $A \cap B$ to 1 and elements of $A \setminus B$ to 0.[6] L_Q, then, just is the set of strings generated by τ from the models in Q.

As an example, consider verifying the sentence 'Most of the dots are blue' against a visual scene with 3 blue dots and 2 yellow (i.e. non-blue) dots. We can represent this scene by a model $\mathcal{M}$ with $M = A = \{d_1, \ldots, d_5\}$ and $B = \{d_1, d_4, d_5\}$. Here, A is the set of dots d_i and B is the set of blue dots. Thus, $A \cap B = B = \{d_1, d_4, d_5\}$ while $A \setminus B = \{d_2, d_3\}$. The encoding described above will generate the string $\tau(\mathcal{M}) = 10011$. The corresponding machine $M_{>1/2}$ has two states: one representing a 'yes' response and one a 'no' response. Inuitively, it pairs off 1s and 0s and returns a 'yes' if and only if there are more 1s than 0s. It processes the string $\tau(\mathcal{M})$ as follows: it pushes the first 1 on to the stack, but pops it off of the stack when encountering the first 0. Then, it pushes the second 0 onto the stack, but pops it off of the stack when encountering the second 1. Then, it pushes the final 1 onto the stack. Because, at the end of processing the string, only 1s are on the stack, the machine accepts the string. This reflects the fact that in $\mathcal{M}$, most of of the dots are in fact blue: $A \cap B$ – the set of blue dots – has a bigger cardinality than $A \setminus B$ – the set of non-blue dots.

In the semantic automata literature thus far, only a single τ has ever been considered.[7] From the psychological perspective, this is surprising: this perspective takes the visual scene to

[4]They used *parity quantifiers* like 'even number of' for reasons that do not concern us here.

[5]See also [Zajenkowski and Szymanik, 2013] for the relationship between intelligence, working memory, executive functions and complexity of quantifiers. [Szymanik, 2016] provides a summary of these experimental results.

[6]That $A \cap B$ and $A \setminus B$ are the only necessary sets follows from the properties of Conservativity and Extensionality. See [van Benthem, 1986].

[7]The only exceptions are when extending the framework to handle more complicated kinds of quantified sentences, e.g. with quantifiers in subject and object position [Steinert-Threlkeld and Icard III., 2013, Steinert-Threlkeld, 201x, McWhirter, 2014].

provide a model $\langle M, A, B\rangle$ with τ embodying how that scene is represented in the mind. Given that different representations of the same situation can effect performance in cognitive tasks, it is natural to suppose that different encodings of finite models might make a relevant difference in quantifier sentence verification. Consider, for instance, the task of verifying 'More than half of the dots yellow' against a visual scene of yellow and blue dots. The τ above will work in every case. If, however, yellow and blue dots are paired together (with all the remaining being of a single color), the agent could make sure that each pair has one dot of each color and that all of the remaining are yellow or blue (corresponding to yes/no answers).

We can model this situation by supposing that the scene makes salient a certain pairing of dots so that the models which are input to the encoding function actually have the form $\langle M, A, B, R\rangle$ where $R \subseteq M \times M$ (here, A is the set of dots and B the set of blue dots). A new mapping τ' can be defined which maps such models into an alphabet containing *pairs* of symbols in addition to 0s and 1s. τ' will map all pairs in R to pairs of symbols in the natural way: if, for instance, $\langle b, a\rangle \in R$ where $b \in A \cap B$ and $a \in A \setminus B$, then $\langle b, a\rangle$ will get mapped to $\langle 1, 0\rangle$. Then, any elements of the model that are not paired will get mapped to 0 or 1 as before. Using this encoding τ', if the only pairs in the encoding are $\langle 1, 0\rangle$ or $\langle 0, 1\rangle$ and all individual symbols are 1s, then the sentence is true. Similarly, if all individual symbols are 0s, it is false. These facts, however, can paradigmatically be checked by a finite-state automaton.

Given that the difference between PDA and FSA-acceptable quantifiers has been shown to correlate very strongly with working memory involvement in verification, the above discussion of encoding suggests that verifying a proportional quantifier sentence against a visual scene in which elements are paired will require less working memory than doing so against a visual scene with randomly scattered elements. Moreover, if distinct but truth-conditionally equivalent quantifiers exhibit different levels of sensitivity to the type of visual scene, this would provide evidence for a constrained relationship between specifications of truth-conditions and verification procedures.

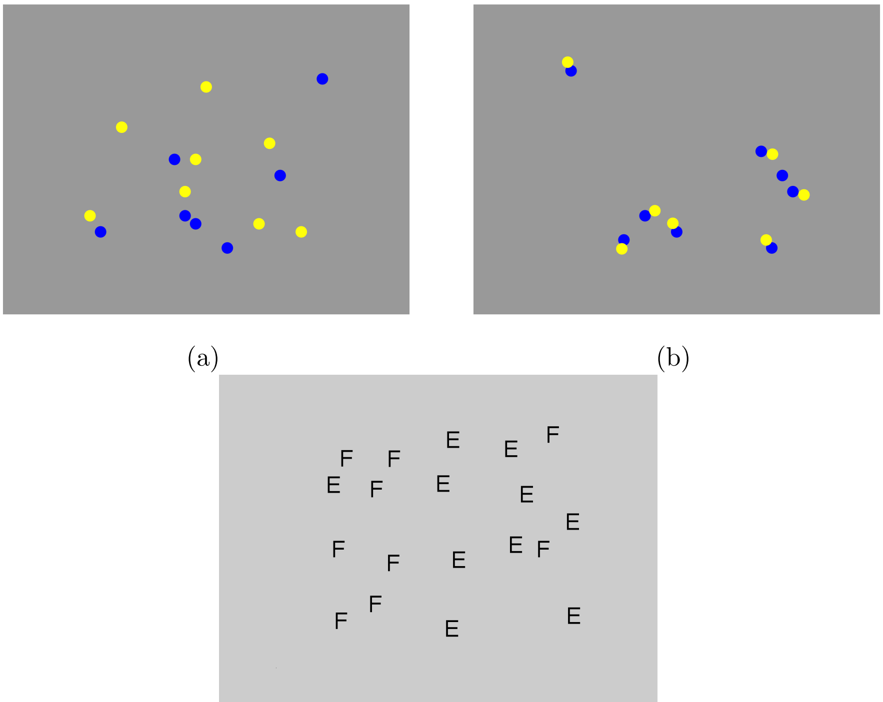
3 Experiment

3.1 Methods

To test this hypotheses, we ran an experiment in which participants had to answer a question while being shown a visual scene containing two types of objects. There were three main conditions, corresponding to the following questions. In parentheses are listed the types of objects in the visual scene.

1. Are more than half of the dots yellow?
   (Yellow and blue dots)
2. Are more than half of the letters 'E'?
   (Characters 'E' and 'F')
3. Are most of the dots yellow?
   (Yellow and blue dots)

The visual scenes came in two types: random, with objects distributed randomly across the image, and paired, with objects presented in pairs consisting of one object of each type. We manipulated the proportions of the two types of objects – 8/7, 9/8 and 10/9 – as well as the correct answer to the question (yes/no). Examples of stimuli demonstrating these variables are given in Figure 1. Participants were randomly assigned to one of the three main conditions and one of the three proportions; the other variables were within-subjects.



(a) (b)

(c)

Figure 1: Example stimuli. (a) Condition: (1) and (3); Proportion: 8/7; Type: random; Answer: yes. (b) Condition: (1) and (3); Proportion: 8/7; Type: paired; Answer: no. (c) Condition: (2); Proportion: 10/9; Type: random; Answer: yes.

To manipulate working memory load, we included a digit recall task. Before the onset of a picture, a participant saw a string of 5 digits for 1500ms. After answering the question against the picture (sentence verification), the participant was presented with one of the digits as a probe and asked to give the digit in the sequence following the probe. In the low memory condition, the same sequence of digits – (0,1,2,3,4) – was used in every trial; in the high memory condition, the digits were randomized [de Fockert et al., 2001]. Participants performed one block in the low memory condition and one in the high condition, with a forced 30 second break in between. Each block consisted of 40 trials (10 in each combination of yes/no and random/paired). The order of the blocks was random as was the order of trials within the block. These main blocks were preceded by a 4-trial training block, after which the participants received feedback on their performance.

3.2 Participants

We recruited participants from Mechanical Turk, all from the United States with HIT approval rate of at least 99%. They were compensated with $1.50. In condition (1), we had 79 participants. We excluded 20 participants who made more than 10 errors on the sentence verification task. The remaining (N = 59) were aged between 20 and 59 (M = 33, SD = 9.9) with 28 male and 31 female. For conditions (2) and (3), we had 60 participants and excluded those who committed more than 30 errors in the verification task. For condition (2), this left us with $N = 54$, 27 male, 27 female, aged between 20 and 69 ($M = 35, SD = 12$). For condition (3), this left us with $N = 57$, 28 male, 27 female, aged between 20 and 68 ($M = 35, SD = 9.6$).

3.3 Results

In both conditions (1) and (2), there is a significant main effect of stimulus type on both the accuracy and reaction time of the sentence verification task. Participants answer faster and more accurately when pictures show paired rather than random dots. Similary, we found main effects of WM condition on both accuracy and RT of the digit recall task in all proportions in both conditions.

To test whether the stimulus type – random or paired – makes a difference in working memory demands, we ran tests to see whether reaction time and accuracy in the *digit recall* task were thereby effected. In particular, we ran a multiple regression of digit recall RT on stimulus type and WM condition as well as a log-likelihood difference test of digit recall accuracy on the same two variables.

Crucially, we found a significant interaction effect of stimulus type and WM condition in digit recall RT of condition (1) in proportion 8/7 ($\chi^2(4) = 4, p = 0.043$). The difference in the RT of the digit task between low and high WM conditions is greater for random pictures (M = 1049 ms, SD = 2193 ms) than for paired pictures (M = 671 ms, SD = 1191 ms). We also find an interaction effect on digit recall accuracy in proportion 8/7 for condition (2) ($\chi^2(1) = 4.19, p < .0407$). The increase in error rate due to the hard WM condition was higher for random pictures (11.25%) than for paired pictures (9%). In condition (2), we also found a trend towards a significant interaction effect on accuracy in the 9/8 proportion ($\chi^2(1) = 3.61, p < .057$). The increase in error rate due to the hard WM condition was higher for random pictures (20%) than for paired pictures (11.18%).

The results for condition (3) – 'most' – are a bit different. While there was a main effect of WM condition on accuracy and RT of digit recall in all proportions, there was no main effect of stimulus type on RT in proportion 9/8 nor on accuracy in proportion 10/9. More importantly, we find no significant interaction effects of stimulus type and WM condition on digit recall RT or accuracy in any proportions. Figure 2 shows the observed main effects and Figure 3 shows the observed interaction effects for all of the conditions.

4 Discussion

Consideration of the role of representation in the semantic automata framework led us to hypothesize that working memory activation in proportional quantifier sentence verification should depend on how a visual scene is presented. In a limited context, we do indeed find such a dependence: in the 8/7 proportion for conditions (1) and (2), the effect of working memory load depended on whether the stimulus type was random or paired. We hypothesize that the effect occurs only at this proportion because participants are more likely to approximate in the larger proportion conditions [Halberda and Feigenson, 2008]. Moreover, that the strongest interaction effects were found in condition (2) – with E/F images – supports this approximation interpretation. Because the two types of objects in the visual scene are so similar to each other, it becomes nearly impossible to approximate. Thus, in the case when approximation is most difficult, we get the strongest interaction effects. This also suggests future manipulations: if we made the 'E's and 'F's colored, perhaps the effect would weaken as approximation becomes available.

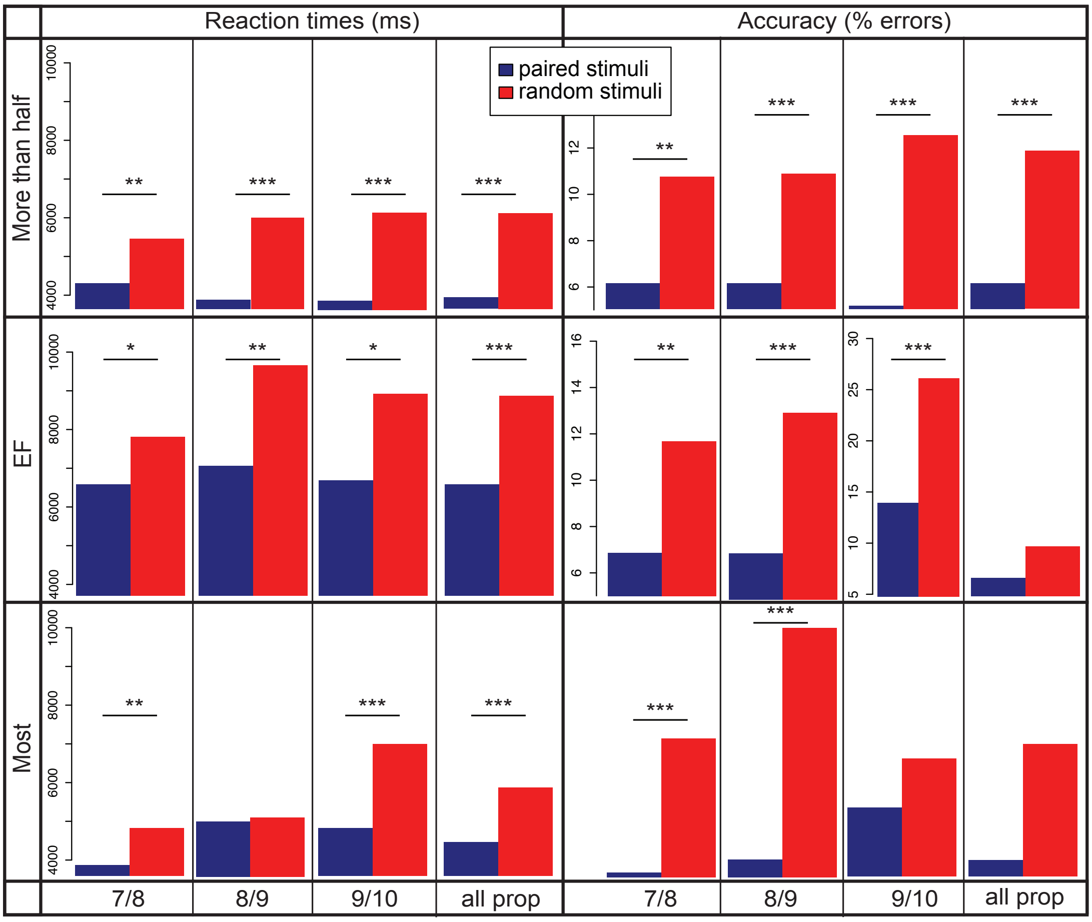


Figure 2: Main effects of stimulus type on the quantifier verification task. Above a column, * means $p < .05$, ** means $p < .01$, and *** means $p < .001$.

4.1 ‘Most’ and ‘More than half’

A very striking feature of the experimental results concerns the very different results between conditions (1) and (3), which differ only in the use of ‘more than half’ in the former and ‘most’ in the latter. In particular, the lack of any interaction effects in the latter suggests that the manipulation of stimulus type does not effect working memory demands for ‘most’ in the way that it does for ‘more than half’. Minimally, this provides evidence for a constrained relationship between specifications of truth-conditions and verification procedures because two truth-conditionally equivalent sentences have very different verification profiles.

Can more about the nature of this constraint be said? [Hackl, 2009] argues that the difference amounts to the following: ‘most’ is a superlative while ‘more than half’ makes explicit size comparisons in terms of proportion. He argues that this difference manifests itself in terms of verification procedures in that ‘most’ will lend itself to a “vote-counting” strategy which resembles pairing colored dots and seeing what colors remain. Interestingly, such an interpretation does not fit well with our data: if ‘most’ lent itself to pairing strategies, one would expect to

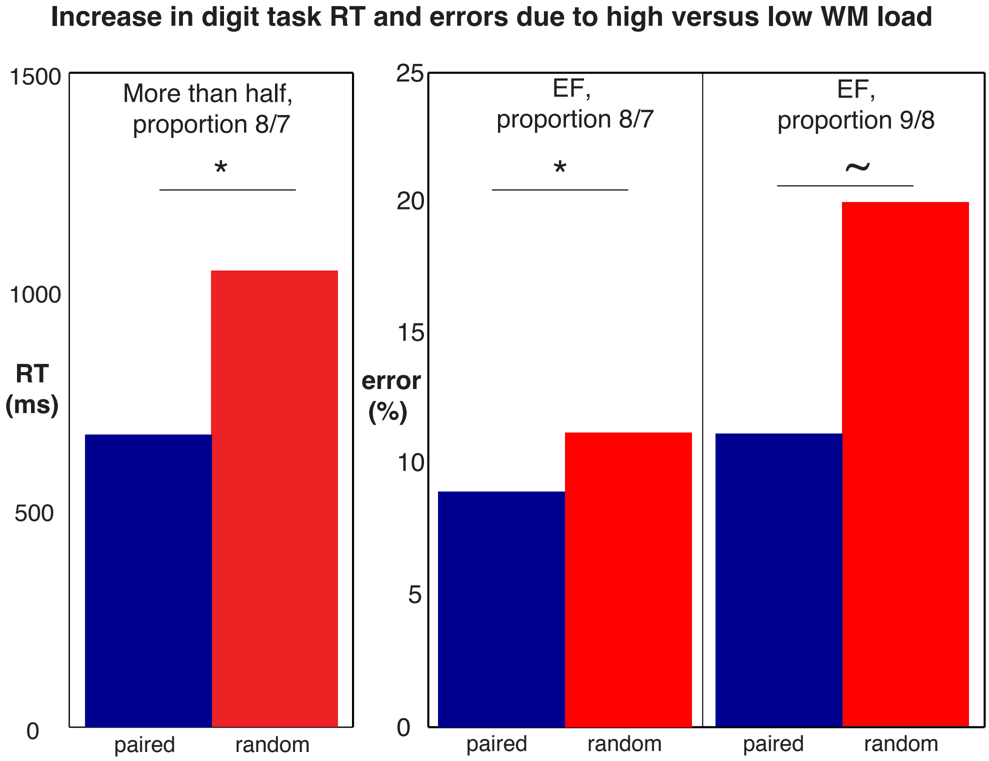


Figure 3: Effects of the interaction of stimulus type and working memory on the digit recall task. Here, $*$ indicates $p < .05$ and $\sim$ indicates a trend toward significance $(0.05 < p < 0.057)$.

see a marked difference in WM demand between random and paired stimuli.[8]

Moreover, [Pietroski et al., 2009] present evidence that 'most' does not favor a vote-counting procedure. They conducted an experiment where participants verified a 'most' sentence against a visual scene of yellow and blue dots which is flashed for 200ms. They find that manipulating the scene between random and paired placement of dots does not effect the accuracy of judgments. Were 'most' to exhibit a bias towards vote-counting verification, one would expect accuracy to improve in paired scenes. At the present, then, we can conclude that 'most' and 'more than half' do indeed verify in their verification behavior; exactly how remains undetermined.

5 Conclusion and Future Directions

In this paper, we have presented novel experimental support for the idea that specifications of truth-conditions for sentences constrain methods of verification. In particular, we used the semantic automata framework to show how different presentations of a visual scene can effect working memory demand. Verifying sentences containing 'more than half', but not those

[8] [Solt, 201x] argues for a very similar difference in the logical forms of 'most' and 'more than half' which places constraints on the types of measurement scale required. She uses this to predict interesting distributional properties and argues that the scale requirements can also account for Hackl's results.

containing 'most', exhibit effects on working memory demand when verified against paired versus random visual scenes. These results present further evidence for a constraint between representations of truth-conditions and verification and a proof-of-concept that working memory can be used to probe such constraints.

Much work, however, remains to be done. On the experimental side, a next step consists in running these experiments using EEG. The memory protocol used in this paper has been shown to be effective in such a context [de Fockert et al., 2001]. As alluded to earlier, more manipulations of our present set-up – color(s) of the items in the stimuli, types of items, the number of proportions considered – could be included. These would provide a fuller picture of the range of data for which a theory of the relationship between truth-conditions and verification needs to account.

On the theoretical side, significant modeling work needs to be done to predict the results here. In particular, how do different quantifiers exhibit a bias for different encodings when visual scenes make them salient? This would help answer why 'more than half', but not 'most', appears to have this bias. At the present moment, the semantic automata framework treats 'most' and 'more than half' equivalently. The experimental results here indicate that this assumption should be relaxed. Moreoever, a fully developed model would need to incorporate the proportion of items in a visual scene as a predictor. The frameworks of resource-rational [Griffiths et al., 2015] or boundedly-rational [Icard, 2014] analysis are likely to be useful in this endeavor.

Acknowledgements Jakub Szymanik was supported by NWO Veni grant 639-021-232.

References

[Barwise and Cooper, 1981] Barwise, J. and Cooper, R. (1981). Generalized Quantifiers and Natural Language. *Linguistics and Philosophy*, 4(2):159–219.

[Braver et al., 1997] Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., and Noll, D. C. (1997). A Parametric Study of Prefrontal Cortex Involvement in Human Working Memory. *NeuroImage*, 5(1):49–62.

[Clark, 2011] Clark, R. (2011). Generalized Quantifiers and Number Sense. *Philosophy Compass*, 6(9):611–621.

[Clark and Grossman, 2007] Clark, R. and Grossman, M. (2007). Number sense and quantifier interpretation. *Topoi*, 26(1):51–62.

[Cummins and Katsos, 2010] Cummins, C. and Katsos, N. (2010). Comparative and Superlative Quantifiers: Pragmatic Effects of Comparison Type. *Journal of Semantics*, 27:271–305.

[de Fockert et al., 2001] de Fockert, J. W., Rees, G., Frith, C. D., and Lavie, N. (2001). The Role of Working Memory in Visual Selective Attention. *Science*, 291(5509):1803–1806.

[Dummett, 1973] Dummett, M. (1973). *Frege: Philosophy of Language*. Harper & Row, New York.

[Geurts et al., 2010] Geurts, B., Katsos, N., Cummins, C., Moons, J., and Noordman, L. (2010). Scalar quantifiers: Logic, acquisition, and processing. *Language and Cognitive Processes*, 25(1):130–148.

[Geurts and Nouwen, 2007] Geurts, B. and Nouwen, R. (2007). At least et al.: The semantics of scalar modifiers. *Language*, 83(3):533–559.

[Griffiths et al., 2015] Griffiths, T. L., Lieder, F., and Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, 7(2):217–229.

[Hackl, 2009] Hackl, M. (2009). On the grammar and processing of proportional quantifiers: *most* versus *more than half*. *Natural Language Semantics*, 17(1):63–98.

[Halberda and Feigenson, 2008] Halberda, J. and Feigenson, L. (2008). Developmental change in the acuity of the "Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, 44(5):1457–1465.

[Horty, 2007] Horty, J. (2007). *Frege on Definitions*. Oxford University Press, Oxford.

[Icard, 2014] Icard, T. (2014). Toward boundedly rational analysis. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*.

[Lidz et al., 2011] Lidz, J., Pietroski, P., Halberda, J., and Hunter, T. (2011). Interface transparency and the psychosemantics of most. *Natural Language Semantics*, 19(3):227–256.

[McMillan et al., 2005] McMillan, C. T., Clark, R., Moore, P., Devita, C., and Grossman, M. (2005). Neural basis for generalized quantifier comprehension. *Neuropsychologia*, 43(12):1729–1737.

[McMillan et al., 2006] McMillan, C. T., Clark, R., Moore, P., and Grossman, M. (2006). Quantifier comprehension in corticobasal degeneration. *Brain and Cognition*, 62(3):250–260.

[McWhirter, 2014] McWhirter, S. (2014). *An Automata-Theoretic Perspective on Polyadic Quantification in Natural Language*. Masters of logic thesis, Universiteit van Amsterdam.

[Moschovakis, 2006] Moschovakis, Y. N. (2006). A logical calculus of meaning and synonymy. *Linguistics and Philosophy*, 29(1):27–89.

[Mostowski, 1998] Mostowski, M. (1998). Computational semantics for monadic quantifiers. *Journal of Applied Non-Classical Logics*, 8(1-2):107–121.

[Pietroski et al., 2009] Pietroski, P., Lidz, J., Hunter, T., and Halberda, J. (2009). The meaning of 'most': Semantics, numerosity and psychology. *Mind and Language*, 24(5):554–585.

[Solt, 201x] Solt, S. (201x). On measurement and quantification: The case of most and more than half. *Language*.

[Steinert-Threlkeld, 201x] Steinert-Threlkeld, S. (201x). Some Properties of Iterated Languages. *Journal of Logic, Language and Information*.

[Steinert-Threlkeld and Icard III., 2013] Steinert-Threlkeld, S. and Icard III., T. F. (2013). Iterating semantic automata. *Linguistics and Philosophy*, 36(2):151–173.

[Suppes, 1980] Suppes, P. (1980). Procedural semantics. In Haller, R. and Grassl, W., editors, *Language, Logic, and Philosophy: Proceedings of the 4th International Wittgenstein Symposium*, pages 27–35. Hölder-Pichler-Tempsy, Vienna.

[Szymanik, 2016] Szymanik, J. (2016). *Quantifiers and Cognition: Logical and Computational Perspectives*. Studies in Linguistics and Philosophy. Springer.

[Szymanik and Zajenkowski, 2010] Szymanik, J. and Zajenkowski, M. (2010). Comprehension of simple quantifiers: empirical evaluation of a computational model. *Cognitive Science*, 34(3):521–532.

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

[Troiani et al., 2009a] Troiani, V., Peelle, J. E., Clark, R., and Grossman, M. (2009a). Is it logical to count on quantifiers? Dissociable neural networks underlying numerical and logical quantifiers. *Neuropsychologia*, 47(1):104–11.

[Troiani et al., 2009b] Troiani, V., Peelle, J. E., McMillan, C., Clark, R., and Grossman, M. (2009b). Magnitude and parity as complementary attributes of quantifier statements. *Neuropsychologia*, 47(12):2684–5.

[van Benthem, 1986] van Benthem, J. (1986). *Essays in Logical Semantics*. D. Reidel Publishing Company, Dordrecht.

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

[Zajenkowski and Szymanik, 2013] Zajenkowski, M. and Szymanik, J. (2013). 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.