

Easy Solutions For a Hard Problem? The Computational Complexity of Reciprocals with Quantificational Antecedents*

Fabian Schlotterbeck and Oliver Bott

Collaborative Research Center 833, University of Tübingen
{fabian.schlotterbeck,oliver.bott}@uni-tuebingen.de

Abstract. The PTIME-Cognition Hypothesis states that cognitive capacities are limited to those functions that are computationally tractable (Frixione 2001). We applied this hypothesis to semantic processing and investigated whether computational complexity affects interpretation preferences of reciprocal sentences with quantificational antecedents, that is sentences of the form *Q dots are (directly) connected to each other*. Depending on the quantifier, some of their interpretations are computationally tractable whereas others are not. We conducted a picture completion experiment and two picture verification experiments and tested whether comprehenders shift from an intractable meaning to a tractable interpretation which would have been dispreferred otherwise. The results suggest that intractable readings are possible, but their verification rapidly exceeds cognitive capacities if it cannot be solved using simple heuristics.

1 Introduction

In natural language semantics sentence meaning is commonly modeled by means of formal logics. Recently, semanticists have started to ask whether their models are cognitively plausible (e.g. Pietroski, Lidz, Hunter & Halberda (2009)). Obviously, for a semantic theory to be cognitively realistic the assumed meanings have to be bounded in computational complexity since they have to be computed in real time by processors with limited resources. In line with this consideration, Frixione (2001) has proposed the *PTIME-Cognition Hypothesis* (PCH) as a general research heuristic for cognitive science: cognitive functions have to be limited to those functions that are computationally tractable (see also van Rooij 2008).

* This research was funded by the German Research Foundation in the Project B1 of SFB 833. The authors would like to thank Fritz Hamm, Janina Radó, Wolfgang Sternefeld, Jakub Szymanik and three anonymous reviewers for valuable comments and discussions on earlier drafts of this paper. We would also like to thank our reviewers and audience at the *International Workshop on Computational Semantics 9* and the *International Conference on Linguistic Evidence 2012* where we have presented parts of the present paper. For their assistance in conducting the reported experiments we would like to give special thanks to Anna Pryslopska and Aysenur Sarcan.

However, computational complexity has hardly received any attention in formal semantics. In this paper we apply the PCH to semantic processing by looking at a particularly interesting test case that has been discussed by Szymanik (2010).

We investigated the interpretation of reciprocal sentences with quantificational antecedents of the form Q (*some N, all N, most N, ...*) *stand in relation R to each other*. What makes these sentences particularly interesting for our purposes is that, depending on the quantificational antecedent, the evaluation of one of their readings is NP-hard whereas other antecedents stay within the realm of PTIME verifiable meanings. We tested whether interpretations are limited to those for which verification is in PTIME (i.e. computationally tractable) as proposed in Szymanik (2010). In particular, we investigated whether comprehenders shift from the preferred interpretation to other – under normal circumstances dispreferred, but computationally tractable – readings in order to avoid having to deal with an intractable verification problem¹. Consider (1).

- (1) a. All of the students know each other.
 b. All of the students followed each other into the room.

Reciprocals like (1) are notoriously ambiguous (Dalrymple, Kanazawa, Kim, McHombo & Peters 1998). The intuitively preferred reading of (1-a) is that any two members of the restriction (= *the students*) have to participate in the reciprocal relation (= *knowing one another*). By contrast, (1-b) states that the students entered the room one after the other, i.e. they can be ordered on a path. Quantified reciprocals like (1) may also exhibit a third reading, i.e. for any member of the restriction there is some other member and these two participate in the reciprocal relation. We dub the first interpretation a complete graph reading, the second a path reading and the third a pair reading. Note the logical dependencies. In fact, the complete graph reading is the logically strongest interpretation entailing the others.

To account for interpretation preferences of sentences like (1) Dalrymple et al. (1998) have put forward the *Strongest Meaning Hypothesis* which proposes that sentences like (1) receive the strongest interpretation consistent with the reciprocal relation. This hypothesis is not undisputed, however, and Kerem, Friedmann & Winter (2010) have presented empirical evidence that it does not hold in general, but that interpreters will choose the most typical interpretation, instead. They refined the original hypothesis accordingly and formulated their *Maximal Typicality Hypothesis*.

The strongest meaning of (1-a), which is presumably also the most typical one, is PTIME verifiable. Interestingly, once we replace the aristotelian quantifier *all* with a proportional quantifier like *most* or a cardinal quantifier like *exactly k*, verification of the strong meaning becomes NP-hard (Szymanik 2010) since it involves solving the CLIQUE problem. In contrast to the *Strongest Meaning Hypothesis*, the PCH therefore predicts that complete graph readings should not be possible for reciprocals with proportional or counting quantifiers as their

¹ A clarification may be in order: When we speak of intractable problems we always refer to NP-hard problems and assume silently that $P \neq NP$.

antecedents. As illustrated in this example, combining the *Strongest Meaning Hypothesis* and the PCH yields specific predictions (and similar predictions can be derived by combining the *Maximal Typicality Hypothesis* with the PCH). We can think of computational complexity as a filter acting on the possible meanings of reciprocal sentences. The effect of this filter should be that the logically strongest meanings is preferred, as long as it is computationally tractable.

The structure of the paper is as follows. In the next section we present a picture completion experiment which was conducted to elicit the preferred interpretation of reciprocals with three different quantificational antecedents. The results show that, in line with our predictions, the interpretation preferences clearly differed between aristotelian and proportional or cardinal quantifier antecedents: we observed only a small proportion of complete graph interpretations in the latter two quantifier types, whereas reciprocals with an aristotelian antecedent were ambiguous. We then present two picture verification experiments that tested whether an intractable complete graph reading was a viable interpretation for proportional and cardinal quantifiers, at all. The results show that potentially intractable complete graph readings are possible as long as they are tested in graphs of limited size. We conclude with a discussion of whether the obtained results require a parameterized version of the PCH (cf. van Rooij & Wareham (2008)) or whether the findings could also be explained if we assume that intractable meanings were approximated using simple heuristics that fail under certain conditions.

2 Picture Completion Experiment

We measured interpretation preferences in a picture completion experiment in which participants had to complete dot pictures corresponding to their preferred interpretation of sentences like (2). As outlined above, the *Strongest Meaning Hypothesis* predicts sentences like (2) to be preferably interpreted with their complete graph meaning (3).

- (2) All/Most/Four of the dots are connected to each other.
 (3) $\exists X \subseteq \text{DOTS} [\text{Q}(\text{DOTS}, X) \wedge \forall x, y \in X (x \neq y \leftarrow \text{CONNECT}(x, y))]$, where Q is ALL, MOST or FOUR.

In combination, the PCH and the SMH predict interpretation differences between the three quantifiers. While the complete graph meaning of *reciprocal all* can be verified in polynomial time, verifying the complete graph interpretation of *reciprocal most* and *reciprocal k* (here: $k = 4$) is NP-hard. By contrast, the weaker readings are computable in polynomial time for all three types of quantifiers. It is thus expected that the choice of the quantifier should affect the preference for complete graph vs. path/pair interpretations: *reciprocal all* should preferably receive a complete graph reading, but *reciprocal most/k* should receive a path or pair reading. The *Maximal Typicality Hypothesis* is hard to apply to these cases because it is unclear what the most typical scenario for *to be connected* should

look like. Most probably, the different readings shouldn't differ in typicality and should hence show a balanced distribution.

2.1 Method

23 German native speakers (mean age 24.3 years; 10 female) received a series of sentences, each paired with a picture of (yet) unconnected dots. Their task was to connect the dots in a way that the resulting picture matched their interpretation of the sentence. We tested German sentences in the following three conditions (*all* vs. *most* vs. *four*).

- (4) Alle (die meisten/vier) Punkte sind miteinander verbunden.
 All (most/four) dots are with-one-other connected.
 All (most/four) dots are connected with each other.

All-sentences were always paired with a picture consisting of four dots, whereas *most* and *four* had pictures with seven dots. In addition to the fifteen experimental trials (five per condition), we included 48 filler sentences. These were of two types. The first half clearly required a complete graph. The second type was only consistent with a path. We constructed four pseudorandomized lists, making sure that two adjacent items were separated by at least two fillers and that each condition was as often preceded by a complete graph filler as it was by a path filler. This was done to prevent participants from getting biased towards either complete graph or path interpretations in any of the conditions. The completed pictures were labeled with respect to the chosen interpretation. We considered a picture to show a complete graph meaning if it satisfied the truth conditions in (3). A picture was taken to be a path reading if a sufficiently large subgraph was connected by a continuous path, but there was no complete graph connecting the nodes. A picture was taken to be a pair reading if the required number of nodes were interconnected, but there was no path connecting them all. Since we didn't find any pair readings, we will just consider the complete graph and path readings in the analysis. Cases that fulfilled neither interpretation were counted as mistakes.

2.2 Results and Discussion

The proportions of complete graph meanings were analyzed using a logit mixed effects model (cf. Jäger (2008)) with *quantifier* as fixed effect and random intercepts of participants and items. Furthermore, we computed three pairwise comparisons: one between *all* and *most*, one between *all* and *four* and one between *most* and *four*.

In the *all*-condition, participants chose complete graph meanings 47.0% of the time. By contrast, in the *most*-condition there were only 22.9% complete graph interpretations among the correct pictures. The number of complete graphs was even lower in the *four*-condition with only 17.4%. The statistical analysis revealed a significant difference between *all* and the other two quantifiers

(*estimate* = -1.87 ; $z = 4.14$; $p < .01$). The pairwise comparisons revealed a significant difference between *all* and *most* (*estimate* = -1.82 ; $z = -3.99$; $p < .01$), a significant difference between *all* and *four* (*estimate* = -3.16 ; $z = -5.51$; $p < .01$), but only a marginally significant difference between *four* and *most* (*estimate* = $.80$; $z = 1.65$; $p = .10$).

The error rates differed between conditions. Participants were 100% correct in the *all*-condition. They made slightly more errors in the *four*-condition which had a mean of 94.8% correct drawings. In the *most*-condition the proportion of correct pictures dropped down to 83.5%. To statistically analyze error rates we computed two pairwise comparisons using Fisher’s exact test. The analysis revealed a significant difference between *all* and *four* ($p < .05$) and a significant difference between *four* and *most* ($p < .01$).

The observed preference for path interpretations, and in particular the very low proportions of complete graph readings for *most* and *four*, matched the predictions of the PCH. Both, *most* and *four* reciprocals, constitute intractable problems and their strong interpretation shouldn’t hence be possible. The error rates provide further support for the PCH. *Most* and *four* led to more errors than *all*. This can be accounted for if we assume that participants were sometimes trying to compute a complete graph interpretation, but due to the complexity of the task did not succeed.

An open question is whether the strong readings of *reciprocal most* and *reciprocal four* are just dispreferred or completely unavailable. This was addressed in a picture verification experiment.

3 Picture Verification Experiment 1

The second experiment tested reciprocals which were presented together with pictures that disambiguated graph from path readings. To achieve clear disambiguation, we had to use different quantifiers than in the previous experiment. This is because the quantifiers were all upward entailing and therefore complete graphs are also consistent with a path reading. In the present experiment, we used reciprocals with *all but one*, *most* and *exactly k*, as in (5). *All but one* and *exactly k* are clearly non-monotone and hence none of the readings entails the others. For *most* it is possible to construct complete graph diagrams in a way that the other readings are ruled out pragmatically if we take its scalar implicature (= *most, but not all*) into account. Crucially, although intuitively more complex than simple *all*, the complete graph reading of *all but one* is PTIME computable. A brute force algorithm to verify reciprocals with *all but one* requires approximately n -times as many steps as an algorithm to verify *all* reciprocals. In order to verify a model of size n , at most the n subsets of size $n - 1$ have to be considered. By contrast, verifying the strong meaning of (5-b,c) is intractable. In particular, *exactly k* is intractable because k is not a constant, but a variable. In the experiment we kept *all but one* constant and presented *exactly k* with the values *exactly three* and *exactly five*.

- (5) (a) Alle bis auf einen / (b) Die meisten / (c) Genau drei (/fünf)
 (a) All but one / (b) The most / (c) Exactly three (/five)
 Punkte sind miteinander verbunden.
 dots are with-one-other connected.
 (a) All but one / (b) Most / (c) Exactly three (/five) dots are connected
 with each other.

We paired these sentences with diagrams disambiguating towards the complete graph or the path reading. Sample diagrams are depicted in the appendix A in Figure 4(a)/(e) and 4(b)/(f), respectively. As for complete graph pictures, the PCH lets us expect lower acceptance rates for (5-b,c) than for (5-a). In order to be able to find out whether the complete graph readings of (5-b/c) are possible at all we paired them with false diagrams which served as baseline controls (see Figure 4(c)/(g)). The controls differed from the complete graph pictures in that a single line was removed from the completely connected subset. If the complete graph reading is possible for (5-b/c), we should observe more “yes, true” judgments in the complete graph condition than in the false controls. As an upper bound we included ambiguous conditions compatible both with complete graph and path conditions (cf. Figure 4 (d)/(i)).

It is possible that people can verify the strong meaning of (5-b,c) given small graphs, but fail to do so for larger ones. Therefore, besides having the three types of quantifiers, another crucial manipulation consisted in the size of the graphs, i.e. the number of vertices. Small graphs always contained four dots (see Fig. 4(a)-(d)) and large graphs consisted of six dots (see Fig. 4(e)-(i)). This way, we also were able to keep the quantifier *all but one* constant and compare it to *exactly k* with variable *k*. *Exactly k* was instantiated as *exactly three* and *exactly five*, respectively. In total, this yielded 24 conditions according to a 3 (*quantifier*) $\times 4$ (*picture type*) $\times 2$ (*graph size*) factorial design.

3.1 Method

We constructed nine items in the 24 conditions and used a latin square design with three lists to make sure that each picture only appeared once for each subject, but that the same picture appeared with all three quantifier types. Each participant provided three judgments per condition resulting in a total of 72 experimental trials. We added 66 filler trials.

36 German native speakers (mean age 26.9 years; 23 female) read reciprocal quantified sentences on a computer screen. After reading the sentence, they had to press a button, the sentence disappeared and a dot picture was presented for which they had to provide a truth value judgment.

3.2 Results

The mean judgments are presented in Figure 1. The proportions of “yes, true” judgments were subjected to two analyses. The lower bound analysis compared

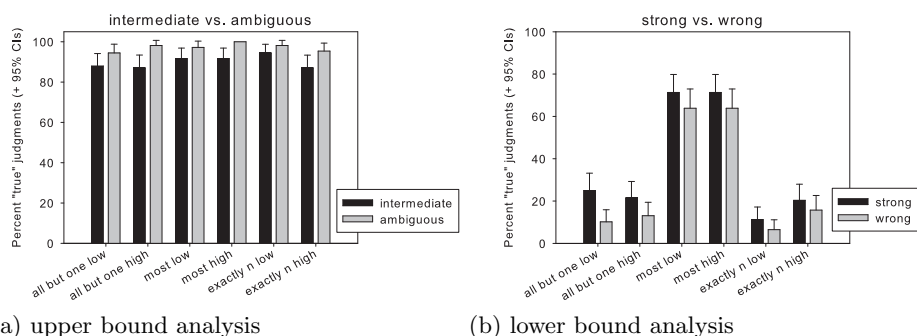


Fig. 1: Mean judgments in Picture Verification Experiment 1 (*low*: pictures with 4 dots; *high*: pictures with 6 dots)

the complete graph conditions with the false baseline controls in order to determine whether complete graph pictures were more often accepted than the latter. The fixed effects of *quantifier* (three levels), *graph size* (two levels) and *picture type* (two levels) were analyzed in a logit mixed effects model with random intercepts of participants and items. Accordingly, upper bound analyses compared the path conditions with the ambiguous conditions.

Upper bound analysis: There was an across the board preference (7.3% on average) of ambiguous pictures over pictures disambiguating towards a path interpretation ($estimate = -2.37$; $z = -2.88$; $p < .01$). No other effects were reliable.

Lower bound analyses: Quantifiers differed with respect to how many positive judgments they received. *Most* received reliably more positive judgments than *exactly k* and *all but one* which led to a significant effect of *quantifier* ($estimate = 3.31$; $z = 8.10$; $p < .01$). There was also a marginally significant effect of *truth* ($estimate = 0.72$; $z = 1.77$; $p = .07$) which was due to slightly higher (7.9%) acceptance of the complete graph pictures as compared to the false baseline controls. No other effects were reliable.

3.3 Discussion

Most behaved rather unexpectedly. Surprisingly, it was rather often accepted in the false baseline controls. The high acceptance rates in these two conditions indicate that participants were canceling its scalar implicature (= *most*, but not *all*) and interpreted it equivalently to the upward monotone *more than half*. This also explains the high acceptance rates of *most* in the complete graph conditions which were, without the implicature, compatible with a path interpretation. Not being able to tell path interpretations apart from complete graph interpretations, we exclude *most* from the following discussion.

Overall, the path reading was strongly preferred and the complete graph reading was hardly available for any quantifier type. However, both the upper

bound and the lower bound analyses provide evidence that the complete graph reading, even though strongly dispreferred, is available to some degree. Both analyses revealed an effect of picture type. Path diagrams were accepted somewhat less often than the ambiguous diagrams and complete graph diagrams were somewhat more often accepted than false diagrams. To our surprise, we didn't find any differences between quantifiers and graph sizes. However, we have to be careful in interpreting these effects because judgments were very close to the floor in the complete graph conditions.

Why did we observe only so few complete graph interpretations even for tractable *all but one* reciprocals? One possible explanation is that readers shifted to path interpretations in all three quantifier conditions because the complete graph readings may have been too complex. In the following we will show that this is not the case, but that the low acceptability of complete graph interpretations was for another reason. In particular, it is due to the reciprocal relation *be connected to each other*. When reconsidering this relation we were not sure whether it had the desired logical properties. It is possible that *to be connected* is preferably interpreted transitively. This could have led to the low number of positive judgments for the complete graph pictures, because, strictly speaking, the complete graph reading would then be false in the complete graph pictures. Note that there was a path connecting all the dots. Thus, assuming transitivity, any dot was pairwise connected to all the others, violating the complete graph reading of *all but one* and *exactly k*.

4 The Reciprocal Relation

To find out whether the reciprocal relation *be connected with each other* allows for a transitive interpretation we conducted a picture verification experiment with non-quantificational reciprocal sentences. We presented 80 participants with the picture in Figure 2 and presented one of the following sentences to each of four subgroups of 20 participants. Each participant provided only one true/false-judgment. This was done to make sure that the data are unbiased.

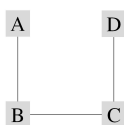


Fig. 2: Diagram used to test for (in)transitivity of the reciprocal relations.

- (6) a. A und D sind miteinander verbunden.
A and D are connected to each other.

- b. A und D sind nicht miteinander verbunden.
A and D are not connected to each other.
- c. A und D sind direkt miteinander verbunden.
A and D are directly connected with each other.
- d. A und D sind nicht direkt miteinander verbunden.
A and D are not directly connected with each other.

The proportion of *yes, true* judgments of sentence (6-a) provides us with an estimate of how easily *be connected with each other* can be interpreted transitively. By contrast, judgments of the negated sentence in (6-b) inform us about how easily the relation *be connected with each other* can be understood intransitively. As a pretest for the next picture verification experiment, we included another relation, *be **directly** connected with each other*, which, according to our intuitions, should only be interpretable intransitively.

4.1 Results and Discussion

The distribution of judgments was as follows. (6-a) was accepted in 80% of the cases, (6-b) was accepted in 42%, (6-c) in 10% and (6-d) in 80% of the cases. Fisher’s exact test revealed that the proportions of “yes, true” judgments in the conditions (6-a), (6-b) and (6-d) were significantly different from 0% ($p < .01$), whereas condition (6-c) didn’t differ significantly from 0% ($p = .49$). With respect to *to be connected* the results show that the relation is in fact ambiguous although there was a preference for the transitive reading as indicated by the higher proportion of “yes, true” judgments of sentence (6-a) than (6-b). The transitive preference might have confined the acceptability of complete graphs in the previous picture verification experiment. Regarding *to be directly connected* there is a strong preference to interpret the reciprocal relation intransitively as revealed by almost no acceptance of (6-c).

5 Picture Verification Experiment 2

In the first picture verification experiment the complete graph interpretation seemed to be possible across quantifiers. If this was really the case it would provide evidence against the PCH. In the present experiment we thus investigated whether this tentative finding can be replicated with a reciprocal relation like *be **directly** connected* which is fully consistent with complete graphs. If the PCH, in its original form, were correct, it would predict complete graphs to be acceptable in tractable *all but one* reciprocals. By contrast, *exactly k* should lead to an interpretive shift and reveal path interpretations, only. If, on the other hand, the findings of the first picture verification experiment could in fact be generalized to intransitive relations, we would expect to find complete graph interpretations of *exactly k*, contra the PCH. Extending this line of reasoning, we might expect the availability of a complete graph reading of *exactly k* reciprocals to be influenced by graph size. In small sized graphs this reading may be fully acceptable,

whereas with increasing size it may well exceed processing capacity. As for *all but one*, both theoretical alternatives predict complete graphs to constitute possible interpretations irrespective of the size of the model.

- (7) (a) Alle bis auf einen / (b) Genau drei (/fünf) Punkte sind
 (a) All but one / (b) Exactly three (/five) dots are
 miteinander verbunden.
 with-one-other connected.
 (a) All but one/ (b) Exactly three (/five) dots are connected with each
 other.

To test these predictions we combined the sentences in (7) with the pictures in Figure 5. The picture conditions were mostly identical to those of the first picture verification experiment. The only difference was that the ambiguous pictures were replaced by two other conditions (see Figure 5(d)/(h)). These were false baseline controls included to find out whether path readings are available with a clearly intransitive relation. The false controls depicted paths that were one edge too short. Altogether this yielded 16 conditions in a $2(\text{quantifier}) \times 2(\text{reading}) \times 2(\text{truth}) \times 2(\text{graph size})$ within subjects design.

Thirty-four native German speakers (mean age: 27.5y, 20 female) took part in the study. Except for the mentioned changes everything was kept identical to the first picture verification experiment.

5.1 Results and Discussion

Mean acceptance rates are depicted in Figure 3. The path and complete graph conditions were analyzed separately using logit mixed effect model analyses. The path reading was generally accepted (true conditions: mean acceptance of 67.7%) and led to almost no errors (false conditions: mean acceptance 4.2%) with both quantifiers and graph sizes. The statistical analysis revealed that only the main effect of *truth* was significant (*estimate* = 5.70; $z = 8.70$; $p < .01$).

The true complete graph diagrams were also accepted across the board (66.3%). Further, true complete graph diagrams were accepted significantly more often than false ones (31%: *estimate* = 1.94; $z = 4.08$; $p < .01$). In the false complete graph conditions there were, however, clear differences between diagrams with four and six dots. In the former conditions participants made relatively few errors (9.9%), whereas error rates increased drastically in the latter conditions (52%). This led to a reliable effect of *graph size* (*estimate* = -6.46; $z = -3.99$; $p < .01$) and a significant interaction of *truth* and *graph size* (*estimate* = 5.23; $z = 3.18$; $p < .01$). Furthermore, the increase in error rates was greater for *exactly k* than for *all but one*. For *exactly k* we observed an increase of 47.0%, whereas there was only a 37.2% increase for *all but one*. Analyzing the false conditions separately we found a significant interaction of *graph size* and *quantifier* (*estimate* = 18.6; $z = 2.04$; $p < .05$).

The relation *to be directly connected* clearly allowed for complete graph readings. As opposed to the predictions of the PCH, this was the case for both *all*

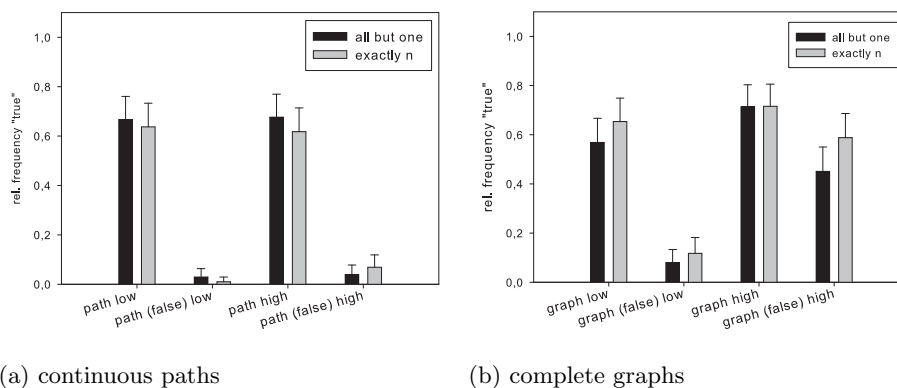


Fig. 3: Mean judgments in Picture Verification Experiment 2 (*low*: pictures with 4 dots; *high*: pictures with 6 dots)

but one and *exactly k*. In the false conditions we did, however, find clear effects of semantic complexity. In these conditions errors drastically increased with the number of vertices. Especially, for the intractable *exactly k* reciprocals error rates increased with the number of dots.

Why did semantic complexity only affect the false conditions? We think that the reason for this may lie in the specific verification strategies participants were able to use in the relevant conditions. In the true complete graph conditions the complete subgraph was visually salient and could be immediately identified. In combination with the fact that the CLIQUE problem is in NP, the true complete graph conditions may, thus, in fact have posed a tractable problem to the participants. As a consequence, only in the false conditions with one edge removed from the graph participants really had to solve an intractable problem. Apparently, this was still possible when the relevant subgraph consisted of only three dots but already started to exceed cognitive capacities when it consisted of five dots.

5.2 Conclusions

We started off with hypotheses that made too strong predictions. Firstly, the linguistic work on reciprocal sentences by Dalrymple et al. (1998) let us expect that reciprocals should be interpreted in their logically strongest meaning. When the antecedent is upward entailing a complete graph interpretation should thus be chosen. Secondly, Szymanik (2010), building upon the PCH, predicted interpretation shifts if reciprocals are intractable due to their quantificational antecedents. Surprisingly, neither of these predictions was fully confirmed. In contrast to the first prediction, complete graph readings were not the default even for tractable reciprocals with an upward entailing antecedent. In the pic-

ture completion study path readings were equally often chosen for reciprocals with *all* as the stronger complete graph reading. At first sight the predictions of the PCH were borne out by the picture completion study. In this experiment the quantificational antecedent affected interpretation preferences according to the predictions of the PCH. When the complete graph reading was intractable it was only rarely chosen. However, in picture verification intractable readings were clearly available to our participants. This provides evidence against the PCH in its most general form.

Still, we did find effects of semantic complexity. Participants had problems to correctly reject pictures not satisfying the complete graph reading with one missing connection. This difficulty increased with the number of dots, especially for intractable *exactly k*. How can these effects be explained? Firstly, it is possible that participants approximated the intractable meaning of *exactly k*, thereby effectively computing tractable functions. These tractable approximations may have worked well in the true complete graph conditions but were inappropriate for the false complete graph controls. We think of specific graphical properties present in the true conditions, e.g. salience of the relevant subgraph which may have simplified the task. Secondly, it is possible that participants were able to compute intractable functions, but only within certain limits, e.g. up to a certain number of elements or in pictures where the relevant subgraph was graphically salient. No matter which of these explanations is correct our data provide an interesting challenge for the PCH. A promising perspective in this respect may be a parameterized version of the PTIME Cognition Hypothesis (cf., for instance, van Rooij & Wareham 2008) which allows us to take into consideration the exact instantiation of the problem. The presence or absence of a perceptually salient group of objects may be a crucial factor for identifying a clique and should, therefore, influence error rates. We plan to explore this in future research.

References

- Dalrymple, Mary, Makoto Kanazawa, Yookyung Kim, Sam McHombo, & Stanley Peters (1998), Reciprocal expressions and the concept of reciprocity, *Linguistics and Philosophy*, 21(2):159–210.
- Frixione, Marcello (2001), Tractable competence, *Minds and Machines*, 11:379–397.
- Jäger, Florian T. (2008), Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models, *Journal of Memory and Language*, 59:434–446.
- Kerem, Nir, Naama Friedmann, & Yoad Winter (2010), Typicality effects and the logic of reciprocity, in *Proceedings of SALT XIX*.
- Pietroski, Paul, Jeffrey Lidz, Tim Hunter, & Justin Halberda (2009), The meaning of ‘most’: semantics, numerosity, and psychology, *Mind & Language*, 24(5):554–585.
- Szymanik, J. (2010), Computational complexity of polyadic lifts of generalized quantifiers in natural language, *Linguistics and Philosophy*, forthcoming.
- van Rooij, Iris (2008), The tractable cognition hypothesis, *Cognitive Science*, 32:939–984.
- van Rooij, Iris & Todd Wareham (2008), Parameterized complexity in cognitive modeling: foundations, applications and opportunities, *The Computer Journal*, 51(3):385–404.

A Sample diagrams

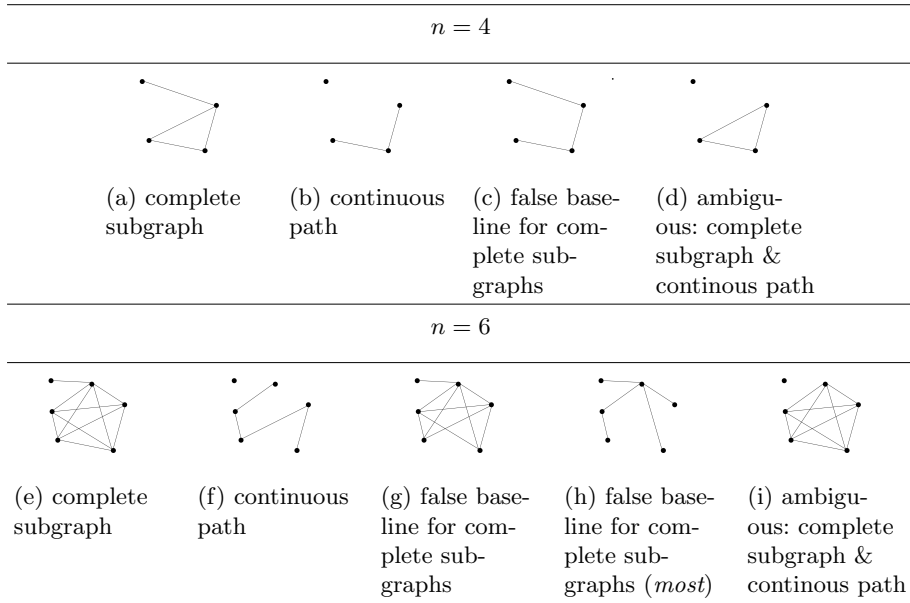


Fig. 4: Sample diagrams presented in Picture Verification Experiment 1. The upper row represents graphs with four dots. Graphs with six dots are represented in the bottom row. The false baseline controls for complete graphs with six dots were slightly different for *all but one* and *exactly five* (g) than for *most* (h). In diagrams like (h), all dots were connected by a path, but in contrast to diagrams like (g) there was no subset containing four or more elements forming a complete graph. This way, for all three quantifiers falsity was due to a small number of missing connections.

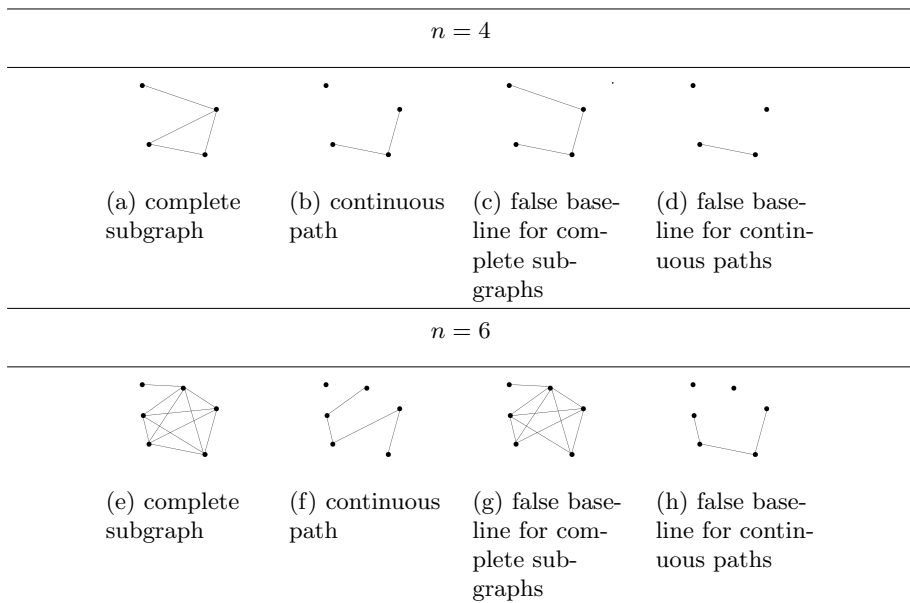


Fig. 5: Sample diagrams presented in the Picture Verification Experiment 2. The upper row represents graphs with four dots. Graphs with six dots are represented in the bottom row.