Deductive reasoning about expressive statements using external graphical representations

Yuri Sato[1], Gem Stapleton[1], Mateja Jamnik[2] and Zohreh Shams[2]
[1]Centre for Secure, Intelligent and Usable Systems, University of Brighton, UK
{y.sato, g.e.stapleton}@brighton.ac.uk
[2]Department of Computer Science and Technology, University of Cambridge, UK
{mateja.jamnik, zohreh.shams}@cl.cam.ac.uk

Abstract

Research in psychology on reasoning has often been restricted to relatively inexpressive statements involving quantifiers. This is limited to situations that typically do not arise in practical settings, such as ontology engineering. In order to provide an analysis of *inference*, we focus on reasoning tasks presented in external graphic representations where statements correspond to those involving multiple quantifiers and unary and binary relations. Our experiment measured participants' performance when reasoning with two notations. The first used topology to convey information via node-link diagrams (i.e. graphs). The second used topological *and* spatial constraints to convey information (Euler diagrams with additional graph-like syntax). We found that topological-spatial representations were more effective than topological representations. Unlike topological-spatial representations, reasoning with topological representations was harder when involving multiple quantifiers and binary relations than single quantifiers and unary relations. These findings are compared to those for sentential reasoning tasks.

Keywords: inference; diagrammatic reasoning; external representation; quantifiers; unary and binary predicates

Introduction

In the literature of psychology of reasoning, categorical syllogisms, e.g., *All B are A; some B are C*; therefore *some C are A*, have been intensively studied (for a survey, see Khemlani & Johnson-Laird 2012). However, it may not be true that categorical syllogisms are most frequently used in our daily life. Non-syllogistic forms of reasoning have attracted particular attention in the study presented in this paper, which encompasses the case of binary verbs (requiring two terms) and multiple quantifiers, such as *All koalas eat only eucalyptuses*. This is achieved through external graphical representations where quantification is implicit in the formed statements; examples will be given later.

Some cognitive studies have been already beyond the traditional framework of categorical syllogisms. Johnson-Laird, Byrne, and Tabossi (1989) dealt with syllogisms involving verbs, for example, from the premises *All A is in the same place as some B; All B is in the same place as all C* to the conclusion *All A is in the same place as all C*. However, the verbs are restricted to spatial binary ones with transitivity or symmetry. The similar restriction was applied to the recent study of Ragni and Sonntag (2012). These studies adopt an approach from the viewpoint of mental model theory, and propose that mental representations for multiply quantified sentences consist simply of alphabets or dots with different shapes representing individuals. However, such simple representations, by way of only using spatial verbs, hardly represent general cases that use binary verbs.

Geurts and van der Slik (2005) used mixed forms of syllogisms with single and double quantifiers; for example, *Most A played against more than two B* and *All B were C* implies *Most A played against more than two C*. They demonstrated the effect of monotonicity profiles of quantifiers, rather than specific mental representations and processes. In addition, non-standard quantifiers, for example, cardinal (numerical) quantifiers such as *more than three* (Kroger et al., 2008), proper nouns such as *a is an A* ("a" is an individual constant) (Politzer & Mercier, 2008; Khemlani, Lotstein & Johnson-Laird, 2014), and proportional quantifiers such as *most* and *few* (Ragni, Singmann, & Steinlein, 2014; Sato & Mineshima, 2016) have been also explored. However, their scopes of the studies were still restricted to the syllogistic form consisting of minor, major, and middle terms. Furthermore, each extended case was explored separately and there have been few comparison between single quantifiers and multiple quantifiers.

By contrast, recent developments in ontology engineering cast a new light on natural language inferences, contributing to the expanding coverage of psychological reasoning tasks. Nguyen et al. (2012) collected deduction patterns (inference tasks), demonstrating the prevalence of a wide variety of forms of reasoning with quantifiers and unary and binary relations. Using novice participants, performance using the deduction patterns as inference tasks was evaluated. It was found that there was a difficulty gap between reasoning with single quantifiers and reasoning with multiple quantifiers, with 76.4% accuracy for (statements involving) single quantifiers and 59.0% for multiple quantifiers. This leads to two important questions. What makes reasoning with multiple quantifiers hard? What kind of cognitive processes of reasoning do people take?

A main concern of these questions is *inference tasks*, which are distinguished from interpretations, but inference necessarily follows the process of interpreting premises. So it is important for the interpretation to be fixed in any way when exploring the nature of human inferences. Indeed, Stenning and van Lambalgen (2001) emphasized the distinction of two kind of reasoning: reasoning *toward* an interpretation of premises and reasoning *from* a fixed interpretation of premisses.[1] Based on the approach of using easy-to-

---

[1]In addition, Stenning and van Lambalgen clearly stated that "If

understand representations (Sato & Mineshima, 2015; Sato et al., 2018), the current study adapts external visual representations, instead of ordinal representations of natural language sentences, to fix the interpretation of premises and provide a fine-grained analysis of inference (see the next section). In particular, we focus on two distinct graphical representations. The first exploits topology to convey information via node-link diagrams (i.e. graphs), hereinafter called a *topological representation*. The second exploits topological *and* spatial constraints to convey information (Euler diagrams with additional graph-like syntax), hereinafter called a *topo-spatial representation*. Importantly, both representations convey information that is semantically rich. They give rise to statements involving either single quantifiers (unary relations) or involving multiple quantifiers along with binary relations.

Task analysis

Topological and topo-spatial representations

As mentioned in the Introduction, human reasoning with quantifiers is necessary in a wide variety of single and multiple quantified forms, in a line with the recent development of ontology engineering. Historically, ontologies had been defined using some kinds of graphs (e.g., Brachman & Schmolze, 1985) since before the formulations of description logics and web ontology languages (W3C OWL Working Group, 2012). As a graph representation for OWL constructs, this study focuses on the Simple Ontology Visualization API: SOVA (Itzik & Reinhartz-Berger, 2014), as shown in Fig.1a-b. Note here that SOVA graphs are fundamentally composed of nodes and links but graphs as full ontology representations need more variant kinds of nodes and links.

In SOVA, nodes are mainly used for classes (i.e., sets), individuals, all things (T), nothing (NT), anonymous classes (A), natural numbers (N), minimum/maximum/exactly cardinalities, negation ($\neg$), intersection ($\cap$), union ($\cup$), binary-verbs with/without universal (all) restriction ($\forall$:) and with/without existential (some) restriction ($\exists$:). In addition, links are used for the relations of isSubclassOf (white-headed arrow), DisjointClasses (double-black-headed reverse arrow), EquivalentClasses (double-black-headed arrow), IntersectionOf (circle-headed arrow to node $\cap$), UnionOf (circle-headed arrow to node $\cup$), ComplementOf (circle-headed arrow to node $\neg$), InstanceOf (line between concept and individual), binary-verbs (black-headed arrow), binary-verbs with universal restriction and "everything" in subject term (white-diamond-headed line), binary-verbs with universal restriction and "everything" in object term (dot-headed line). SOVA graphs, whilst more complex than those as in Hartley and Barnden (1997), but are still composed of topological relationships in nodes and links, giving rise to semantic meaning of expressive statements involving quantifiers.

As well as topological relations, spatial relations – such as inclusion and exclusion – are available. We focus on concept diagrams (Stapleton et al., 2017; 2018) as a topo-spatial

the interpretation is not fixed, what is one actually testing? (p.291)".

representation of ontologies, as shown in Fig.1c-d. The basic idea of concept diagrams is that Euler diagrams and graphs (nodes and links) are merged; for a similar approach, see Harel (1988), although that system is not expressive enough to fully represent ontologies. In concept diagrams, rectangles represent all things, named (labelled) curves represent classes, unnamed curves are used for anonymous classes, dots for individuals, shading illustrates the absence of things. Also numbers and inequality symbols ($\geq n$, $\leq n$) place cardinality constraints (e.g., ≥ 1 means at least one thing), solid arrows with labels (binary-verbs with universal restriction), dashed arrows with labels (binary-verbs with existential restriction), and dashed arrows with inequalities (binary-verbs with number restrictions) are used. Furthermore, relations are divided into two types: set-theoretical one and binary-verb. Set-theoretical relations (e.g., SubclassOf, DisjointClasses, EquivalentClasses, IntersectionOf, UnionOf, ComplementOf, InstanceOf) are expressed by spatial (inclusion and exclusion) relations of the corresponding syntactic objects. Binary-verb relations are expressed by arrows from source objects to target objects.

Reasoning with multiple quantifiers in topological and topo-spatial representations

We can make inferences using topological representations of SOVA graphs and using topo-spatial representations of concept diagrams. In such diagrammatic reasoning tasks, one is asked to judge whether the diagram transformations from premises to a conclusion are valid. Here, the premise diagrams are true and they are transformed into the conclusion diagram. If the conclusion diagram is true, given the information in the premise diagrams, the transformation is valid. Otherwise the transformation is not valid.

Fig.1 shows examples of tasks and Fig.2 illustrates possible intermediate representations generated by merging premise diagrams. In each figure, (a) and (b) are SOVA cases (c) and (d) are concept diagram (Euler or Venn diagram) cases. Cases (a) and (c) give rise to statements about sets and unary relations involving single quantifiers, hereinafter called the 'single quantifier' case, and are translated as *Everything is a darfellan & No grippli is a darfellan; therefore, nothing is a grippli*. Cases (b) and (d) give rise to the statements about sets and binary relations involving multiple (double) quantifiers, hereinafter called the 'multiple quantifier' case, and are translated as *Every daemonfey is related to at least one thing in both axani and phoera under 'isGuidedBy' & No axani is a phoera; therefore, nothing is a daemonfey*.

Here we can observe that, in reasoning with single quantifiers, SOVA and concept (Euler/Venn) diagrams are similar in that unary relations are expressed by one basic component: SOVA links two nodes and concept diagrams use spatial relations. In SOVA, expressing binary verbs (between quantifies) requires *multiple* arrows among nodes.[2] Thus, transforming

---

[2]Stenning (2002, Chap.2) pointed out that graph (node-link) representations are essentially same as sentential representations in that
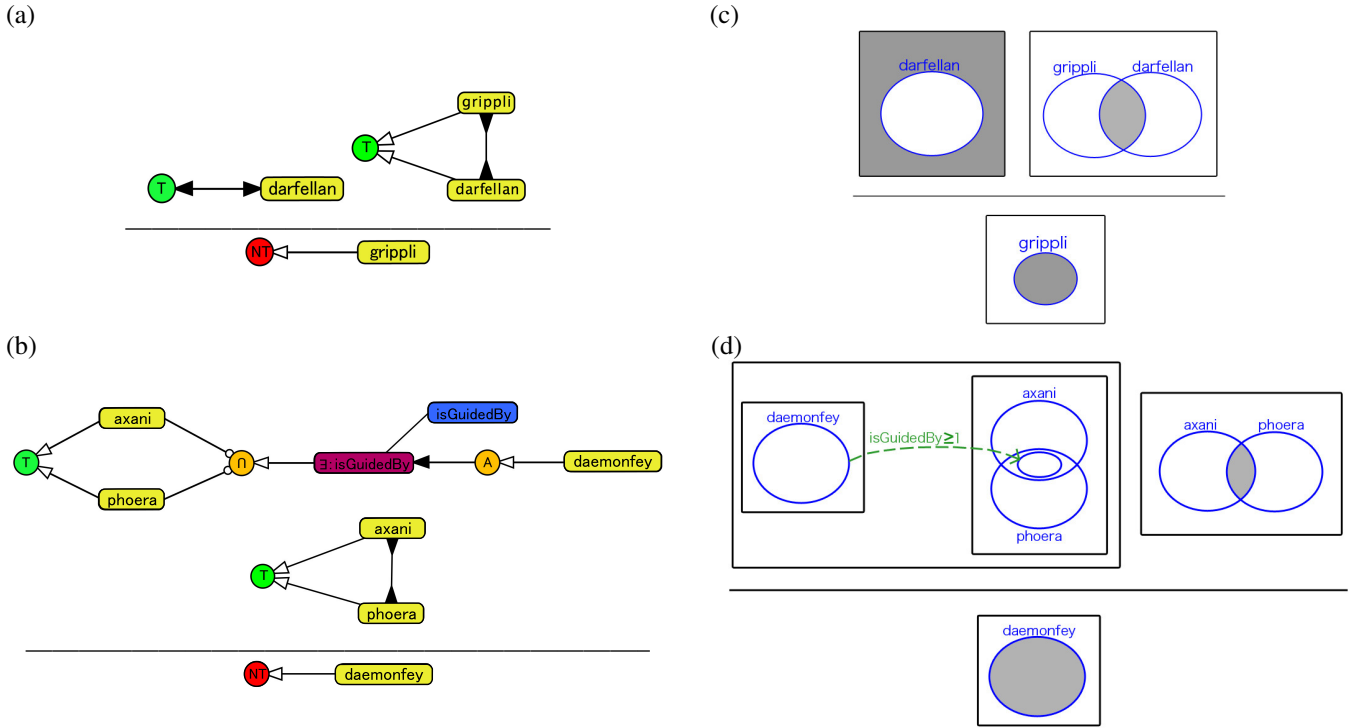
Fig. 1: Task examples. Two premises and one conclusion are divided by a line: (a) topological representations in single quantifier case; (b) topological representations in multiple quantifier case; (c) topo-spatial representations in single quantifier case; (d) topo-spatial representations in multiple quantifier case. (a) and (c) mean *Everything is a darfellan & No grippli is a darfellan; therefore, nothing is a grippli*. (b) and (d) mean *Every daemonfey is related to at least one thing in both axani and phoera under 'isGuidedBy' & No axani is a phoera; therefore, nothing is a daemonfey.*
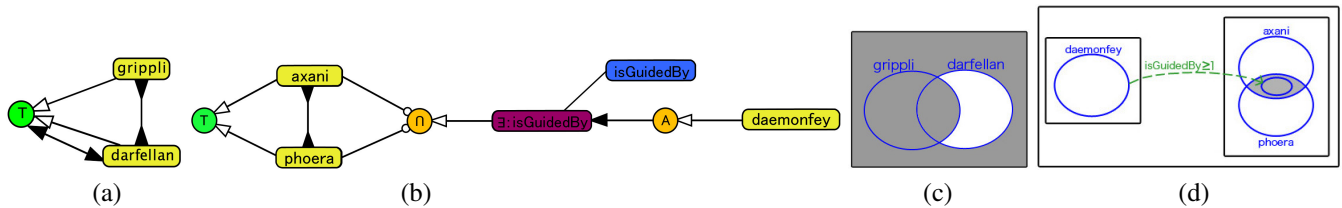

Fig. 2: Possible intermediate representations by merging premise diagrams, corresponding to the cases (a)–(d) in Fig.1

SOVA in reasoning with multiple quantifiers is expected to require much more cognitive effort than reasoning with single quantifiers. On the other hand, in concept diagrams, binary verbs are expressed by one arrow in which the source and target are directly specified (since unary relations are expressed by spatial constraints instead of arrows). Thus, reasoning with multiple quantifiers may not require much more effort than reasoning with single quantifiers.

We therefore make four predictions. (1) reasoning with single quantifiers: there is no significant difference between the two representations. (2) reasoning with multiple quantifiers: topo-spatial representations are more effective than topological representations. (3) reasoning with topological representations: multiple quantifiers are harder than single quantifiers. (4) reasoning with topo-spatial representations:

there is no significant difference between single and multiple quantifiers.

Experiment

Method

Participants Fourty-five undergraduate students from classes on elementary computer science in the University of Brighton were recruited. The mean age was 22.53 ($SD = 5.92$) with a range of $18 - 48$ years. All participants gave informed consent and were paid for their participation. The experiment method was approved by the CEM School Research Ethics Panel of University of Brighton. None had any prior training of ontology engineering or syllogistic logic. One participant gave up on the way, and their data was excluded. Participants were divided into two groups: the topological group ($N = 19$) and the topo-spatial group

---

there is no direct correspondence between syntax and semantics.

($N = 25$), following a between group design.

Materials The participants in each group were presented with premise graphs/diagrams and a conclusion graph/diagrams (such as Fig.1). Participants were asked to answer the question whether the graph/diagram transformations from premises to a conclusion were valid. As shown in table 1 (appendix), we presented 20 items in total, out of which 10 items consisted of validly transformed diagrams and 10 items included invalidly transformed diagrams. The valid 10 items were selected from the medium difficulty patterns in Nguyen et al. (2012). The invalid items were created from the valid ones with minimum changes. Furthermore, the tasks were divided into singly quantified cases and multiply quantified cases. The tasks were presented in one of three random orders and as a paper-and-pencil test. There was no time limit for completing the tasks, although the approximate time (30 minutes) for taking the experiment was instructed.

Procedures All participants were collected in a room. First, the participants were provided with three pages of instructions on the basic meaning of SOVA graphs or concept diagrams, but not on particular rules to solve inference tasks. Second, a pretest to check whether they understood the instructions correctly was conducted; they were asked to choose, from a list of three possibilities, the sentence corresponding to a given graph/diagram (the potential difference of familiarities of representations is reduced since both groups received substantial instruction and underwent practice trials: cf. Sato & Mineshima, 2015). Third, the participants were provided with one page of instruction on the meaning of valid transformation (entailment), with two examples of graphs/diagrams: one was valid and one was invalid.[3] After the instruction phase, the participants were asked to solve the main tasks on reasoning.

Results and Discussion

The data of the participants who made mistakes in more than two items (out of five) of the pretest were removed. In the following analysis, 3 out of 19 in the topological group, and 4 out of 25 participants in the topo-spatial group were removed.

Figure 3 shows the average accuracy rates of inference tasks in the two groups; for each task result, see table 1 in the appendix. The data was subjected to two-way ANOVA for a mixed design. There was a significant main effect of notation factor (i.e., topological vs. topo-apatial), $F = 4.435$, $p = 0.042$. There was a significant main effect of factor involving quantifiers (i.e., single vs. multiple quantifiers), $F = 4.712$, $p = 0.037$. There was no significant interaction effect, $F = 0.426$, $p = 0.518$. Regarding the total 20 items, the accuracy rates in the topo-spatial group were significantly higher than those in the topological group: 58.4% for the topological group and 69.1% for the topo-spatial group.

A post-hoc test by Bonferroni's method was conducted. Regarding the reasoning with single quantifiers, there was no significant difference: 63.8% for the topological group and 71.9% for the topo-spatial group $F = 1.834$, $p = 0.071$. This is consistent with our first prediction. Regarding the reasoning with multiple quantifiers, the accuracy rates in the topo-spatial group were significantly higher than those in the topological group: 53.1% for the topological group and 66.2% for the topo-spatial group $F = 2.938$, $p = 0.005$. This supports our second prediction. In the topological group, reasoning with multiple quantifiers were significantly harder than reasoning with single quantifiers, $F = 2.849$, $p = 0.007$. This conforms to our third prediction. In the topo-spatial group, there was no significant difference between single quantifiers and multiple quantifiers, $F = 1.532$, $p = 0.134$, consistent with our fourth prediction.

In each comparison between valid and invalid items, there was no significant difference at the threshold of 5% in two-tailed t-tests. (i) 58.1% for valid items vs. 58.8% for invalid items in the topological group ($t = -0.085$); 71.9 for valid items vs. 66.2% for invalid items in the topo-spatial group ($t = 1.059$) (ii) 66.3% for valid items with single quantifiers vs. 61.3% for invalid items with single quantifiers in the topological group ($t = 0.516$); 79.0% for valid items with single quantifiers vs.64.8% for invalid items with single quantifiers in the topo-spatial group ($t = 1.878$) (iii) 50.0% for valid items with multiple quantifiers vs. 56.3% for invalid items with multiple quantifiers in the topological group ($t = -0.682$); 64.8% for valid items with multiple quantifiers vs. 67.8% for invalid items with multiple quantifiers in the topo-spatial group ($t = -0.370$). This shows that the accuracy performance is not different between valid and invalid items.

General discussion

In summary, our experiment suggests that topo-spatial representations, such as concept diagrams, can be more effective than topological representations, such as SOVA, in reasoning tasks containing multiply quantified information. In topological representations, reasoning with multiple quantifiers was harder than reasoning with single quantifiers. But in topo-spatial representations, there was no significant difference between these cases. That is, in the topo-spatial case, the difficulty of the inference task did not increase when richer information was conveyed by the premises.

Regarding the performance difference between single and multiple quantifiers, it is noted that the tendency of reasoning from topological representations is similar to those of sentential reasoning, reported in Nguyen et al. (2012) as mentioned in the Introduction. This finding sheds light on the fact that mental representations elicited from sentences contain simple components only, rather than the hybrid components which realize some efficient way to represent tasks. Indeed, whether model-like representations based on linked data points (Johnson-Laird et al., 1989; Ragni & Sonntag,
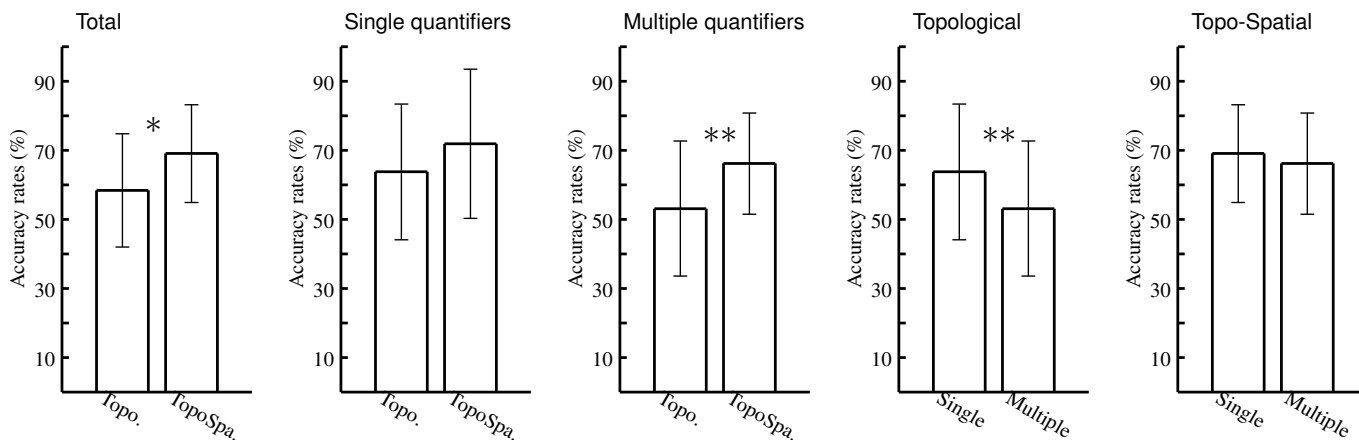
Fig. 3: Average accuracy rates of inference tasks in the topological group and the topo-spatial group. Error bars represent one standard deviation. $*$ refers to a significant difference at $p < 0.05$. $**$ refers to $p < 0.01$.

2012; Greene, 1992) or syntactic representations corresponding to parsing trees/graphs of sentences (Braine, 1998), mental representations to express multiple quantifiers are assumed to be more complex than those of single quantifiers. Of course, the complexity can be reduced, for example, by using additional inventions such as set-theoretical and spatial notions, rather than linking of each simple component. However, the inventions seem to be rarely used spontaneously in people not-trained in logic and mathematics. Accordingly, Bucciarelli and Johnson-Laird (1999) pointed out that the use of Euler circles is a sophisticated method provided by school education and is not natural as naive people's mental representations of quantified assertions (p.296).

Can the result that topo-spatial representations were effective in reasoning with multiple quantifiers provide some implication to cognitive science? The findings of effective expression of tasks can contribute to *task naturalization* in the psychology of reasoning (Politzer, Bosc-Miné & Sander, 2017). If the aim of the current research is measuring people's actual logical (not puzzle solving) capability, tasks involving inference, as opposed to just interpretation, should be set for participants [4]. Our experiment implies that the way to express single quantifiers as spatial relations and multiple quantifiers as topological relations, rather than both types expressed as topological relations, is natural for the cognitive task of reasoning with multiple quantifiers.

Furthermore, our findings in inference or entailment judgment are in contrast to those of consistency checking (Sato et al., 2017), where topological representations were more effective than topo-spatial representations. Consistency checking is a kind of logical reasoning in which people are asked to answer the question of whether the meaning of a diagram was contradictory. The contrast between these empirical findings suggests that there are two distinct cognitive process under-

lying logical reasoning from external diagrams. One is a pattern matching strategy, especially based on *syntactic* forms of representations. While patterns of conclusions entailed from premises are unlimited (but in the case of syllogisms, there are some restrictions), there are certain (common) patterns of inconsistency that are exhibited by statements. In the tasks of consistency checking, then, the strategy to syntactically match patterns to target representations can be reasonable. This strategy is suitable for notations which are expressed in a uniform way. Thus, this suggests to us why topological representations were superior than topo-spatial representations in consistency checking tasks. Another strategy is a more *semantic*. As shown in Fig.1, the conclusion such as *Nothing is an A* requires reasoners to generate some new objects which cannot be found only in syntactic manipulations of representations. Such processes are available only when reasoners correctly understand semantic meanings of given representations. This semantic process might be suitable for the topo-spatial representations.

In this study, we dealt with deductive inference patterns beyond the forms of syllogisms, but, of course, all of them were not covered. For example, we did not handle relations between verbs, e.g., 'bought' isInverseOf 'wasSoldBy', which is a common style of premise in ontology engineering. So these should be analyzed in a next step. Through further extended studies, the nature of human reasoning in general would be explored.

Appendix

Lists of experimental tasks and their results are shown in Table 1. #01–10 are valid items; #11–20 are invalid items. #01–05/#11–15 are relevant to single quantifiers only; #06–10/#16–20 are relevant to multiple quantifiers.

---

[4]For example, the errors caused by the misinterpretation of premises should be blocked here, as discussed in the Introduction.

Table 1: Lists of experimental tasks and their results.

No	Premises ⇒ Conclusion	Topo.%	TopoSpa.%
01.	(Every *A* is a *B*) & (no *A* is a *B*) ⇒ (nothing is an *A*)	62.5	80.1
02.	(No *A* is a *B*) & (every *C* is an *A*) & (every *D* is a *B*) ⇒ (no *C* is a *D*)	75.0	85.7
03.	(Every *A* is a (every *B* is a *C*)) & (every *B* is a *C*) ⇒ (every *A* is a *C*)	87.5	85.7
04.	(Everything is a *B*) & (no *A* is a *B*) ⇒ (nothing is an *A*)	56.3	66.7
05.	(Every *A* is a *B*) & (every *A* is non-*B*) ⇒ (nothing is an *A*)	50.0	76.2
06.	(Every *A* is related to at least one *B* under *R*) & (Everything that something is related to under *R* is a *C*) ⇒ (every *C* is related to at least one thing in both *B* and *C*)	87.5	80.1
07.	(Every *A* is related to at least one thing in both *B* and *C* under *R*) & (no *B* is a *C*) ⇒ (nothing is an *A*)	25.0	85.7
08.	(Every *A* is related to at least three things in *B* under *R*) & (every *A* is related to at most one *B* under *R*) ⇒ (nothing is an *A*)	43.8	52.4
09.	(Every *A* is related to at least one *B* under *R*) & (every *B* is nothing) ⇒ (nothing is an *A*)	62.5	57.1
10.	(Every *A* is related to at least four things in *B* under *R*) & (each thing is related to at most one thing under *R*) ⇒ (nothing is an *A*)	31.3	47.6
11.	(Every *B* is a *A*) & (no *A* is a *B*) ⇒ (nothing is an *A*)	75.0	71.4
12.	(No *A* is a *B*) & (every *C* is an *A*) & (every *B* is a *D*) ⇒ (no *C* is a *D*)	37.5	52.4
13.	(Every *A* is a (every *B* is a *C*)) & (every *B* is a *C*) ⇒ (every *A* is a *B*)	43.8	47.6
14.	(Everything is a *B*) & (every *A* is a *B*) ⇒ (nothing is an *A*)	62.5	66.7
15.	(Every *A* is a *B*) & (every non-*B* is an *A*) ⇒ (nothing is an *A*)	87.5	85.7
16.	(Every *A* is related to at least one *B* under *R*) & (Everything that something is related to under *R* is a *C*) ⇒ (Everything that something is related to under *R* is both a *C* and a non-*B*)	43.8	38.1
17.	(Every *A* is related to at least one thing in either *B*, *C*, or both, under *R*) & (no *B* is a *C*) ⇒ (nothing is an *A*)	62.5	85.7
18.	(Every *A* is related to at least 1 thing in *B* under *R*) & (every *A* is related to at most three things in *B* under *R*) ⇒ (every *A* is nothing)	56.3	66.7
19.	(Every *A* is related to at least one *B* under *R*) & (every *A* is nothing) ⇒ (nothing is a *B*)	62.5	66.7
20.	(Every *A* is related to at least one *B* under *R*) & (each thing is related to at most one thing under *R*) ⇒ (nothing is an *A*)	56.3	81.0

References

Brachman, R. J., & Schmolze, J. G. (1985). An overview of the KL-ONE knowledge representation system. *Cognitive Science*, 9, 171–216.

Braine, M. D. S. (1998). Steps toward a mental-predicate logic. In M.D.S. Braine & D.P.O'Brien (Eds.), *Mental Logic* (pp. 273–331). Mahwah, NJ: Lawrence Erlbaum.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science*, 23, 247–303.

Geurts, B. & van Der Slik, F. (2005). Monotonicity and processing load. *Journal of Semantics*, 22, 97–117.

Greene, S. B. (1992). Multiple explanations for multiply quantified sentences: Are multiple models necessary? *Psychological Review*, 99, 184–187.

Harel, D. (1988). On visual formalisms. *Communications of the ACM*, 31, 514–530.

Hartley, R. T., & Barnden, J. A. (1997). Semantic networks: visualizations of knowledge. *Trends in Cognitive Sciences*, 1, 169–175.

Itzik, N., & Reinhartz-Berger, I. (2014). SOVA - A tool for semantic and ontological variability analysis. In *Proc. CAiSE 2014 Forum*, CEUR vol 1164 (pp. 177–184).

Johnson-Laird, P. N., Byrne, R. M., & Tabossi, P. (1989). Reasoning by model: The case of multiple quantification. *Psychological Review*, 96, 658–673.

Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, 138, 427–457.

Khemlani, S., Lotstein, M., & Johnson-Laird, P. (2014). A mental model theory of set membership. In *Proc. 36th Annu. Conference of the Cognitive Science Society* (pp. 2489–2494). Austin, TX: Cognitive Science Society.

Kroger, J. K., Nystrom, L. E., Cohen, J. D., & Johnson-Laird, P. N. (2008). Distinct neural substrates for deductive and mathematical processing. *Brain Research*, 1243, 86–103.

Nguyen, T. A. T., Power, R., Piwek, P., & Williams, S. (2012). Measuring the understandability of deduction rules for OWL. In *Proc. 1st Internat. Worksh. on Debugging Ontologies and Ontology Mappings* (pp. 1–12), Linköping University Press.

Politzer, G., Bosc-Miné, C., & Sander, E. (2017). Preadolescents solve natural syllogisms proficiently. *Cognitive Science*, 41, 1031–1061.

Politzer, G., & Mercier, H. (2008). Solving categorical syllogisms with singular premises. *Thinking & Reasoning*, 14, 434–454.

Ragni, M., Singmann, H., & Steinlein, E. M. (2014). Theory comparison for generalized quantifiers. In *Proc. 36th Annu. Conference of the Cognitive Science Society* (pp. 1222–1227). Austin, TX: Cognitive Science Society.

Ragni, M., & Sonntag, T. (2012). Preferences and illusions in quantified spatial relational reasoning. *Cognitive Processing*, 13, 289–292.

Sato, Y. & Mineshima, K. (2015). How diagrams can support syllogistic reasoning: an experimental study. *Journal of Logic, Language and Information*, 24, 409–455.

Sato, Y., & Mineshima, K. (2016). Human reasoning with proportional quantifiers and its support by diagrams. In *Proc. Diagrams 2016, LNCS 9781* (pp. 123–138). Springer.

Sato, Y., Stapleton, G., Jamnik, M., Shams, Z., & Blake, A. (2017). How Network-based and set-based visualizations aid consistency checking in ontologies. In *Proc. 10th International Symposium on Visual Information Communication and Interaction* (pp. 137–141). ACM.

Sato, Y., Sugimoto, Y., & Ueda, K. (2018). Real objects can impede conditional reasoning but augmented objects do not. *Cognitive Science*, 42, 691–707.

Stapleton, G., Compton, M., & Howse, J. (2017). Visualizing OWL 2 using diagrams. In *Proc. 2017 IEEE Symposium on Visual Languages and Human-Centric Computing* (pp. 245–253). IEEE Computer Society Press.

Stapleton, G., Delaney, A., Compton, M., & Chapman, P. (2018). Visualizing $\mathcal{ALC}$ using concept diagrams. In *Proc. 5th Internat. Worksh. on Graph Structures for Knowledge Representation and Reasoning, LNAI 10775* (pp. 99–117), Springer.

Stenning, K. (2002). *Seeing Reason: Image and Language in Learning to Think*. Oxford: Oxford University Press.

Stenning, K., & van Lambalgen, M. (2001). Semantics as a foundation for psychology: A case study of Wason's selection task. *Journal of Logic, Language and Information*, 10, 273–317.

W3C OWL Working Group (2012). *OWL 2 Web Ontology Language*. http://www.w3.org/TR/owl2-overview/