

Visual Logics Help People: An Evaluation of Diagrammatic, Textual and Symbolic Notations

Eisa Alharbi, John Howse, Gem Stapleton, Ali Hamie and Anestis Touloumis
University of Brighton, Brighton, UK
{e.alharbi2,John.Howse,g.e.stapleton,a.a.hamie,a.touloumis}@brighton.ac.uk

Abstract—Our aim is to provide empirical evidence that diagrammatic logics are more effective than symbolic and textual logics in allowing people to better *understand* information. Ontologies provide an important focus for such an empirical study: people need to understand the axioms of which ontologies comprise. A between-groups study compared six frequently-used axiom types using the (textual) Manchester OWL Syntax (MOS), (symbolic) description logic (DL) and concept diagrams. Concept diagrams yielded significantly better task performance than DL for all six, and MOS for four, axiom types. MOS outperformed concept diagrams for just one axiom type and DL for only three axiom types. Thus diagrams could ensure ontologies are developed more robustly.

I. INTRODUCTION

A major motivation for the development of diagrammatic logics is the belief that they help people better understand and reason about information as compared to symbolic or textual representations. Theories cited to support this belief include that of free rides [1], [2], generalized to the theory of observational advantages [3].

One major application area for logic is that of ontology engineering. Symbolic description logic and stylized textual OWL are commonly used notations for ontology engineering. Some ontology engineers will be well placed to use these notations, but domain experts may find them difficult to understand. This potential difficulty is not mere conjecture. The inaccessibility of DL and OWL is recognized by Rector et al. [4]: “Understanding the logical meaning of any description logic ... is difficult for most people, and OWL-DL is no exception.” Warren et al. [5] back this insight up with an empirical study of the most commonly used OWL axiom types. They found that “despite training, users are prone to certain misconceptions” and they observed that “one-third of [participants] commented on the value of drawing a diagram.” They further recognize that existing visualizations are “chiefly aimed at viewing the structure of the [ontology] ... rather than the more cognitively difficult features of Description Logics” and they identify that concept diagrams [6] are the exception. Therefore, it seems both important and timely to establish whether user *understanding* of ontologies can be improved through the use of concept diagrams.

This paper sets out to establish, for six common axiom types, which of concept diagrams, OWL (in the Manchester OWL Syntax form presented in Protégé [7]), and description logic allow people to perform tasks more effectively. We

measure effectiveness in terms of accuracy and time performance, with data collected via an empirical study. The tasks we consider in this paper are related to *understanding* axioms. We present a brief overview of concept diagrams in section II. In section III, we describe the design and execution of the empirical study. The results are presented in section IV and discussed in section V. Details of the questions and training material used in the study, along with the raw data collected and its statistical analysis, can be found at <https://sites.google.com/site/eisamalharbi/understandingontologies>.

II. CONCEPT DIAGRAMS

We present a brief overview of concept diagrams with reference to features from our empirical study. Closed curves represent sets. The spatial relationships between the curves assert set-theoretic relationships. For example, the diagram in figure 1 expresses ‘All Sea Elves are Wild and Aquatic’ and figure 2 expresses ‘No Goblin is Tall or Friendly’. The bounding box represents the class Thing (the universal set).

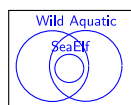


Fig. 1. Subsumption.

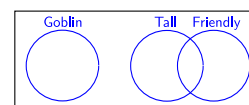


Fig. 2. Disjointness.

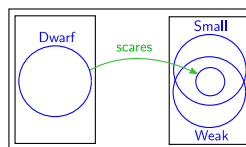


Fig. 3. All values from.

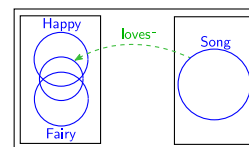


Fig. 4. Some values from.

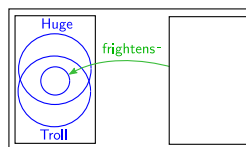


Fig. 5. Domain.

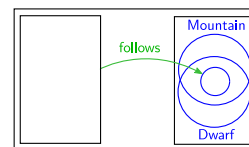
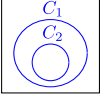
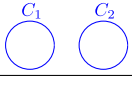
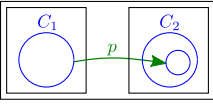
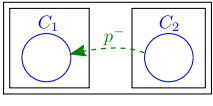
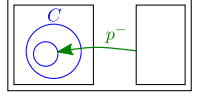
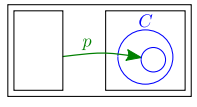


Fig. 6. Range.

Binary relations, called *properties* in ontology engineering, are represented by arrows. The source and target of the arrow, between them, express a property restriction. Arrows can be of two types: solid and dashed. We illustrate their meaning by

TABLE I
REPRESENTING AXIOMS

Axiom Type	Concept Diagram	MOS	DL	Answer Choices
Subsumption		$C_2 \text{ SubClassOf } C_1$	$C_2 \sqsubseteq C_1$	All C_2 are C_1 All C_1 are C_2 At least one C_2 is a C_1 No C_2 is a C_1
Disjointness		$C_1 \text{ DisjointWith } C_2$	$C_1 \sqcap C_2 \sqsubseteq \perp$	No C_2 is a C_1 All C_2 are C_1 At least one C_2 is a C_1 At least one C_2 is not a C_1
All VF		$C_1 \text{ SubClassOf } p \text{ only } C_2$	$C_1 \sqsubseteq \forall p.C_2$	C_1 p only C_2 C_1 p at least one C_2 At least one C_1 p C_2 At least one C_1 p only C_2
Some VF		$C_1 \text{ SubClassOf } p \text{ some } C_2$	$C_1 \sqsubseteq \exists p.C_2$	C_1 p at least one C_2 C_1 p only C_2 At least one C_1 p C_2 At least one C_1 p only C_2
Domain		$p \text{ Domain } C$	$\exists p.\top \sqsubseteq C$	Only C p Things Not only C p Things C p at least one Thing At least one C p Things
Range		$p \text{ Range } C$	$\top \sqsubseteq \forall p.C$	Things p only C Things p at least one C At least one Thing p only C At least one Thing p C

example. Figure 3 represents three classes: Dwarf, Small and Weak. By using multiple bounding boxes, it does not assert any relationship between Dwarf (in one bounding box) and the other two classes (in the other bounding box). The solid arrow is labelled *scares* and is sourced on Dwarf. The unlabelled curve at its target represents the set of things that are scared by dwarfs. Since this curve is contained by both Small and Weak, it thus represents a subset of both these two sets. The diagram expresses that ‘Dwarfs scare only things that are Small and Weak’, corresponding to an OWL axiom involving ‘all values from’. Figure 4 uses a dashed arrow, labelled by *loves⁻*, the inverse of *loves*. The target of the arrow represents a *subset* of the set of things that songs *loves⁻*. That is, things in the set represented by the unlabelled curved love at least one song. From this, we can see the meaning of the diagram: ‘Happy Fairies love at least one Song’, corresponding to an OWL axiom involving ‘some values from’. Figure 5 expresses ‘Only Huge Trolls frighten things’, corresponding to a domain axiom. Figure 6 expresses a range axiom: ‘Things follow only Mountain Dwarfs’. Concept diagrams can make a much richer variety of statements, see [8].

III. EMPIRICAL STUDY DESIGN

In order to determine which notation – concept diagrams, MOS, or DL – most effectively helps people understand ontologies, we focused on six commonly occurring axiom types: *class subsumption*, *class disjointness*, *all values from*, *some values from*, *domain*, and *range*. We tested the following hypothesis for each of the six axiom types: concept diagrams allow people to understand axioms significantly

more effectively than MOS which, in turn, is significantly more effective than DL. We measured understanding through accuracy and time performance data, with accuracy being the primary indicator. We designed an empirical study that required participants to answer multiple-choice questions. The axioms presented information about mythical creatures to give some context to the questions. A between-group design was used, with participants being exposed to one of the three notations, to reduce any learning effect.

We used a standard way of representing each of the six axiom types with each of the three notations, see table I; the concept diagrams chosen for these axiom types are based on the empirical results in [9], [10]. The participants were provided with four multiple-choice options, exactly one of which was correct, indicated in table I. The multiple-choice options were chosen by considering common mistakes made by novice ontology users [4], [5], [11], [12], [13]. For each question, the multiple-choices were presented in random order which was the same for each participant. Each axiom type was used in three questions, each time representing different information, giving a total of 18 questions for each participant.

A pilot study was conducted to test the experiment design, software used to display the questions, and the data collection process. Fifteen participants (3F, 12M, ages 19-30) took part in the pilot study, five per group. No changes were required after the pilot study. A further 75 participants (15F, 60M, ages 18-45) took part in the main study, 25 in each group. All participants were students from the University of Brighton and were randomly allocated to groups. The experiment had three phases: paper-based training, software-based training,

and the main study. The 18 questions were presented in a random order, generated separately for each participant, to reduce learning effect. The experiment was performed in the university’s usability laboratory. Participants performed the experiment individually and were offered a canteen voucher for their time spent in the study.

IV. RESULTS

For the statistical analysis, we employed a Generalized Estimating Equations (GEE) methodology [14]. The basic statistical tool utilized to analyze the data for accuracy, and separately for time, is a GEE-based regression model using the axiom type and the group (CD, DL, or MOS) as explanatory variables that are linearly connected with the logarithms of the odds of a correct answer. The significance of the explanatory variables and their interaction will be assessed to determine whether they affect the probability of answering a question more accurately or more quickly. The analysis was performed using the statistical software R and functions from the R package *geepack* [15].

We collected data from from 90 participants (30 per group); as no changes were made after the pilot study, we included these data in the statistical analysis. Each participant answered 18 questions giving a total of 1620 observations, 270 for each axiom type and 540 for each group. For time data, we analyzed only those times arising from correct answers, leaving 1251 observations. The results for both accuracy and time are given in table II. We analyse the results for Class Subsumption in detail and then summarize the results for all the axiom types.

Class Subsumption Analysis The estimated odds of providing a correct answer in the CD group is 4.0000 times that in the DL group with a 95% confidence interval (CI) of (1.0809, 14.8028) and a p -value of 0.0379. Therefore CDs significantly outperform DL. Similarly, CDs significantly outperform MOS but there is no significant difference between DL and MOS. The estimated time for providing a correct answer in the CD group is 0.8652 times that in the DL group with a 95% CI of (0.7729, 0.9686) and a p -value of 0.0119. Therefore CDs significantly outperform DL. Similarly, CDs significantly outperform MOS but there is no significant difference between DL and MOS. In summary, CDs support significantly more accurate and significantly faster performance than both MOS and DL. There was no significance difference in accuracy or time between DL and MOS.

Summary of Analysis The main results are:

- CDs allow people to understand all six axiom types significantly more effectively than DL.
- CDs allow people to understand four axiom types – subsumption, disjointness, all values from, and range – significantly more effectively than MOS.
- MOS allows people to understand one axiom type – domain – significantly more effectively than CDs.

TABLE II
RESULTS

Subsumption				
Accuracy	Estimate	95% LB	95% UB	p-value
CD vs DL	4.0000	1.0809	14.8028	0.0379
CD vs MOS	6.0000	1.6384	21.9728	0.0068
DL vs MOS	1.5000	0.5526	4.0720	0.4262
Timing	Estimate	95% LB	95% UB	p-value
CD vs DL	0.8652	0.7729	0.9686	0.0119
CD vs MOS	0.8027	0.7102	0.9073	0.0004
DL vs MOS	0.9280	0.8160	1.0540	0.2510
Disjointness				
Accuracy	Estimate	95% LB	95% UB	p-value
CD vs DL	4.2609	0.8642	21.0074	0.0750
CD vs MOS	2.5789	0.4796	13.8692	0.2697
DL vs MOS	0.6053	0.1970	1.8593	0.3806
Timing	Estimate	95% LB	95% UB	p-value
CD vs DL	0.6190	0.5630	0.6810	0.0000
CD vs MOS	0.7740	0.7060	0.8490	0.0000
DL vs MOS	1.2500	1.1300	1.3800	0.0000
All Values From				
Accuracy	Estimate	95% LB	95% UB	p-value
CD vs DL	68.0588	8.4276	549.6261	0.0001
CD vs MOS	56.6364	7.0057	457.8666	0.0002
DL vs MOS	0.8322	0.3419	2.0254	0.6856
Timing	Estimate	95% LB	95% UB	p-value
CD vs DL	0.7510	0.6670	0.8440	0.0000
CD vs MOS	0.7890	0.7140	0.8720	0.0000
DL vs MOS	1.0510	0.9230	1.1960	0.4530
Some Values From				
Accuracy	Estimate	95% LB	95% UB	p-value
CD vs DL	2.3780	0.9001	6.2824	0.0805
CD vs MOS	2.4912	0.9310	6.6659	0.0691
DL vs MOS	1.0476	0.4319	2.5413	0.9181
Timing	Estimate	95% LB	95% UB	p-value
CD vs DL	0.8512	0.7600	0.9534	0.0054
CD vs MOS	0.9480	0.8480	1.0610	0.3530
DL vs MOS	1.1140	0.9832	1.2622	0.0902
Domain				
Accuracy	Estimate	95% LB	95% UB	p-value
CD vs DL	5.1250	1.5338	17.1243	0.0079
CD vs MOS	1.7305	0.4483	6.6797	0.4261
DL vs MOS	0.3377	0.1066	1.0696	0.0649
Timing	Estimate	95% LB	95% UB	p-value
CD vs DL	0.9170	0.8170	1.0300	0.1440
CD vs MOS	1.1274	1.0251	1.2399	0.0135
DL vs MOS	1.2294	1.0958	1.3792	0.0004
Range				
Accuracy	Estimate	95% LB	95% UB	p-value
CD vs DL	62.1321	7.6441	505.0130	0.0001
CD vs MOS	27.0870	3.2651	224.7076	0.0022
DL vs MOS	0.4360	0.1665	1.1417	0.0910
Timing	Estimate	95% LB	95% UB	p-value
CD vs DL	0.7480	0.6760	0.8290	0.0000
CD vs MOS	0.8350	0.7680	0.9080	0.0000
DL vs MOS	1.1153	1.0122	1.2289	0.0274

- MOS allows people to understand three axiom types – disjointness, domain, and range – significantly more effectively than DL.

V. DISCUSSION

The results of the empirical study clearly show that axioms presented in concept diagrams are significantly better understood than those presented in symbolic DL for all axiom types and MOS for most axiom types. Among the

theories proposed to explain the better performance of people using diagrammatic logics observed here is that of well-matchedness [16]. A notation is well-matched to meaning if its semantic relationships are matched by its syntactic relationships. For example, the semantics of *subsumption*, that C_2 is a subset of C_1 , is matched by the curve C_1 containing the curve C_2 in its diagrammatic representation. The concept diagrams used in the study for *subsumption*, *disjointness*, *all values from* and *range* axioms are particularly well-matched to meaning and participants performed well with them, relative to the other notations. The other two axiom types, *some values from* and *domain*, gave rise to the highest error rates for concept diagrams. This could be explained by the use of the cognitively more difficult inverse property p^- , which is not well-matched to meaning. In the case of *some values from*, though, it is notable that both MOS and DL have particularly high error rates too, indicating that this axiom is cognitively difficult.

As described earlier, the multiple-choice options were selected by considering common mistakes made by ontology users. Table III summarizes the types of error the participants made for each axiom type for each notation. It is well-known that *all values from* and *some values from* cause problems for people [4], [5], [11], [12], [13]. Considering the participants in the DL group, 20 answers (22%) incorrectly interpreted an *all values from* axiom as *some values from*. In addition, 18 answers (20%) incorrectly interpreted a *some values from* axiom as *all values from*; the corresponding figures for those participants in the MOS group are 13 (14%) and 7 (8%), respectively. However, the corresponding error data for the concept diagrams group are 1 (1%) and 6 (7%), respectively; this may indicate that the diagrammatic representation goes some way to alleviating this misunderstanding.

Some values from has been shown to be interpreted incorrectly in other research; for example, considering the pizza ontology, many users interpreted “Pizza hasTopping MozzarellaTopping” as “some pizzas have toppings that are mozzarella topping”, rather than the correct reading, “all pizzas have toppings that are some mozzarella topping” [11]. The *some values from* section of table III shows that many participants made this (mis)interpretation: 13 such errors for CDs, 19 for MOS and 14 for DL. A possible reason for this kind of misunderstanding may be that participants associated ‘at least one’ with the wrong class. The classic logical error that ‘all’ implies ‘some’ could account for the errors enumerated in those rows of table III where the incorrect response starts “At least one”. It is noticeable that participants who used concept diagrams very rarely made this kind of error; only 21 errors of this form were made by participants using concept diagrams, compared with 90 made by participants using MOS and 91 by those using DL. This may well indicate that the visualization provided by concept diagrams mitigates against users making this kind of error.

TABLE III
RESPONSES

Subsumption Responses	CD	MOS	DL
Correct: All C_2 are C_1	84	63	70
Total Incorrect	6	27	20
Incorrect: All C_1 are C_2	5	10	6
Incorrect: At least one C_2 is a C_1	1	11	2
Incorrect: No C_2 is a C_1	0	6	12
Disjointness Responses	CD	MOS	DL
Correct: No C_2 is a C_1	84	76	69
Total Incorrect	6	14	21
Incorrect: All C_2 are C_1	3	3	8
Incorrect: At least one C_2 is a C_1	1	8	7
Incorrect: At least one C_2 is not a C_1	2	3	6
All Values From Responses	CD	MOS	DL
Correct: C_1 p only C_2	89	55	51
Total Incorrect	1	35	38
Incorrect: C_1 p at least one C_2	1	13	20
Incorrect: At least one C_1 p C_2	0	16	8
Incorrect: At least one C_1 p only C_2	0	6	10
Some Values From Responses	CD	MOS	DL
Correct: C_1 p at least one C_2	71	54	55
Total Incorrect	19	36	35
Incorrect: C_1 p only C_2	6	7	18
Incorrect: At least one C_1 p C_2	13	19	14
Incorrect: At least one C_1 p only C_2	0	10	3
Domain Responses	CD	MOS	DL
Correct: Only C p Things	82	77	60
Total Incorrect	8	13	30
Incorrect: Not only C p Things	4	6	10
Incorrect: C p at least one Thing	3	4	10
Incorrect: At least one C p Things	1	3	10
Range Responses	CD	MOS	DL
Correct: Things p only C	89	69	53
Total Incorrect	1	21	37
Incorrect: Things p at least one C	1	13	18
Incorrect: At least one Thing p only C	0	2	5
Incorrect: At least one Thing p C	0	6	14

VI. CONCLUSION

This paper set out to provide empirical evidence that diagrams can be more effective than symbolic and textual logics. We considered notations used for ontology engineering, the Manchester OWL Syntax and DL, and compared them to concept diagrams to determine which notation most effectively helped people understand ontologies. By focusing on six commonly occurring axioms, we established the following:

- 1) CDs were more effective than DL for all axiom types.
- 2) CDs were more effective than MOS for four of the six axiom types.
- 3) MOS was more effective than DL for three axiom types.

The results of this study provide evidence that concept diagrams help people understand axioms and that they could play an important role in supporting ontology engineering. As such, future work will consider including concept diagrams in tools such as Protégé, to allow people to take advantage of their cognitive benefits.

Acknowledgements This research was partially funded by a Leverhulme Trust Research Project Grant (RPG-2016-082) for the project entitled Accessible Reasoning with Diagrams.

REFERENCES

- [1] A. Shimojima, "Inferential and expressive capacities of graphical representations: Survey and some generalizations," in *3rd International Conference on the Theory and Application of Diagrams*. Springer, 2004, pp. 18–21.
- [2] —, *Semantic Properties of Diagrams and Their Cognitive Potentials*. Stanford, CA, USA: CSLI Publications, 2015.
- [3] G. Stapleton, M. Jamnik, and A. Shimojima, "Effective representation of information: Generalizing free rides," in *9th International Conference on the Theory and Application of Diagrams*. Springer, 2016, pp. 296–299.
- [4] A. Rector, N. Drummond, M. Horridge, J. Rogers, H. Knublauch, R. Stevens, H. Wang, and C. Wroe, "OWL Pizzas: Practical Experience of Teaching OWL-DL," in *Engineering Knowledge in the Age of the Semantic Web*, ser. LNCS, E. Motta, N. Shadbolt, A. Stutt, and N. Gibbins, Eds. Springer, 2004, vol. 3257, ch. 5, pp. 63–81.
- [5] P. Warren, P. Mulholland, T. Collins, and E. Motta, "The Usability of Description Logics," in *The Semantic Web: Trends and Challenges*, ser. LNCS, vol. 8465. Springer, 2014, pp. 550–564.
- [6] J. Howse, G. Stapleton, K. Taylor, and P. Chapman, "Visualizing ontologies: A case study," in *International Semantic Web Conference*. Springer, 2011, pp. 257–272.
- [7] "Protégé," <http://protege.stanford.edu/>, accessed March 2017.
- [8] G. Stapleton, J. Howse, P. Chapman, A. Delaney, J. Burton, and I. Oliver, "Formalizing Concept Diagrams," in *Visual Languages and Computing*. Knowledge Systems Institute, 2013, pp. 182–187.
- [9] E. Alharbi, J. Howse, G. Stapleton, and A. Hamie, "Evaluating diagrammatic patterns for ontology engineering," in *9th International Conference on the Theory and Application of Diagrams*. Springer, 2016, pp. 51–66.
- [10] T. Hou, P. Chapman, and A. Blake, "Antipattern comprehension: an empirical evaluation," in *International Conference on Formal Ontology in Information Systems*. IOS Press, 2016, pp. 211–224.
- [11] M. Horridge, N. Drummond, J. Goodwin, A. Rector, R. Stevens, and H. Wang, "The Manchester OWL syntax," in *OWLed*, vol. 216, 2006.
- [12] A. Rector, N. Drummond, M. Horridge, J. Rogers, H. Knublauch, R. Stevens, H. Wang, and C. Wroe, "Designing user interfaces to minimise common errors in ontology development: The co-ode and hyontuse projects," in *UK e-Science All Hands Meeting*, vol. 2004, 2004, pp. 493–499.
- [13] R. Schwitter and M. Tilbrook, "Controlled natural language meets the semantic web," in *Proc. Australasian Language Technology Workshop*, vol. 2004, 2004, pp. 55–62.
- [14] K. Liang and S. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, pp. 13–22, 1986.
- [15] J. Yan and J. Fine, "Estimating equations for association structures," *Statistics in Medicine*, vol. 23, pp. 859–880, 2004.
- [16] C. Gurr, "Effective diagrammatic communication: Syntactic, semantic and pragmatic issues," *Journal of Visual Languages and Computing*, vol. 10, no. 4, pp. 317–342, 1999.