Human Reasoning with Proportional Quantifiers and Its Support by Diagrams

Yuri Sato[1(✉)] and Koji Mineshima[2]

[1] Interfaculty Initiative in Information Studies, The University of Tokyo, Tokyo, Japan
satoyuri0@gmail.com

[2] Center for Simulation Sciences, Ochanomizu University, Tokyo, Japan
mineshima.koji@ocha.ac.jp

Abstract. In this paper, we study the cognitive effectiveness of diagrammatic reasoning with proportional quantifiers such as *most*. We first examine how Euler-style diagrams can represent syllogistic reasoning with proportional quantifiers, building on previous work on diagrams for the so-called plurative syllogism (Rescher and Gallagher, 1965). We then conduct an experiment to compare performances on syllogistic reasoning tasks of two groups: those who use only linguistic material (two sentential premises and one conclusion) and those who are also given Euler diagrams corresponding to the two premises. Our experiment showed that (a) in both groups, the speed and accuracy of syllogistic reasoning tasks with proportional quantifiers like *most* were worse than those with standard first-order quantifiers such as *all* and *no*, and (b) in both standard and non-standard (proportional) syllogisms, speed and accuracy for the group provided with diagrams were significantly better than the group provided only with sentential premises. These results suggest that syllogistic reasoning with proportional quantifiers like *most* is cognitively complex, yet can be effectively supported by Euler diagrams that represent the proportionality relationships between sets in a suitable way.

Keywords: Euler diagrams · Proportional quantifiers · Reasoning · Logic and cognition

1 Introduction

Euler diagrams have been used to represent various set-theoretical properties and relations. We can distinguish three types:

(i) The basic inclusion and exclusion relations between sets, as represented by sentences like *All A are B* and *No A are B*;
(ii) The proportionality relationship between sets, as represented by sentences like *Most A are B* and *More than half of A are B*;
(iii) The cardinality of sets, represented by sentences like *Three A are B*, *More than three A are B*, and *Less than three A are B*.

M. Jamnik et al. (Eds.): Diagrams 2016, LNAI 9781, pp. 123–138, 2016.
DOI: 10.1007/978-3-319-42333-3_10

In previous cognitive studies on Euler diagrams [16,17], empirical evidence has been found in support of the effectiveness of diagrams with respect to (i). However, whether and how diagrams can also be effective in representing and reasoning about (ii) and (iii) still remains to be explored. In this paper, we will focus on (ii), i.e., the proportionality relationship between sets, and examine the cognitive effectiveness of diagrams to represent proportional quantification as expressed by a sentence like *Most A are B*.

As we will review in Sect. 2, logical properties of proportional quantifiers like *most* have been the focus of recent research on quantification in logic and linguistics. In logical and cognitive studies on diagrams, however, although Euler and Venn diagrams are widely used to represent and reasoning with quantified statements, little has been discussed about how they can represent proportional quantification and how effective they can be in actual reasoning. This paper is a first step to fill this gap.

The structure of this paper is as follows. In Sect. 2, we present backgrounds on proportional quantifiers in logical, computational and cognitive studies. In Sect. 3, we analyze diagrammatic representations (Sect. 3.1) and diagrammatic inferences (Sect. 3.2) for proportional quantifiers in order to generate predictions (Sect. 3.3). In Sect. 4, we report the results of an experiment comparing participants' performance in solving proportional syllogism with and without diagrams. Section 5 concludes the paper with a summary and discusses some directions for future work.

2 Background on Proportional Quantifiers

We will start with a brief overview of the logical, computational and cognitive properties of proportional quantifiers. This provides the necessary background information about the main issue in this paper.

Logic. Natural languages use many expressions of quantification. Among them, quantifiers such as *all*, *some* and *no* can be represented within first-order logic, using the *unary* quantifiers $\forall$ and $\exists$ and other logical connectives. Thus, the sentence *All A are B* is represented as $\forall x(Ax \rightarrow Bx)$, the sentence *Some A are B* as $\exists x(Ax \wedge Bx)$, and the sentence *No A are B* as $\forall x(Ax \rightarrow \neg Bx)$.

In contrast, it is known that quantifiers that denote the proportionality relation between sets are not definable within first-order logic [3]. A typical example is the quantifier *most*, where *Most A are B* is usually analyzed to mean *More than half of A are B*, symbolized as $|\mathbf{A} \cap \mathbf{B}| > |\mathbf{A}|/2$, or equivalently, $|\mathbf{A} \cap \mathbf{B}| > |\mathbf{A} - \mathbf{B}|$. As these paraphrases show, the quantifier *most* essentially denotes the binary relation between sets, which is not reducible to a standard unary quantification (i.e., a property of a set). Throughout this paper, we call a non-first-order quantifier like *most* a *proportional quantifier*, in contrast to a standard first-order quantifier like *all* and *no*.

Computation. Quantifier interpretations in terms of generalized quantifier theory have also been analyzed from computational perspectives. In a seminal work, van Benthem [4] uses automata to model semantic computing of quantified sentences. For example, in the standard quantified sentence of the form *All A are B*, the machine reads the states of objects. If the object is *B*, the transition to an accepting state occurs; otherwise, the transition to a rejecting state occurs. Thus, the machine need not memorize data at each process, and the machine's system can be realized by a simple finite automaton. By contrast, a proportional quantified sentence of the form *Most A are B* can only be modeled by push-down automata. In the case of proportional quantifiers, if the result of reading the states of objects is equal to the information in the top stack, the result is stored in a pushdown stack; otherwise, information in the stack is removed. Thus, a memory device (pushdown stack) is needed to give a computational modeling of proportional quantifiers. The resulting system is realized not by a finite automaton but by a push-down automaton. In this sense, the difference between proportional and non-proportional quantification also appears in the semantic automata approach to quantifiers.

Cognition. In cognitive psychology, it has been discussed whether the logical and computational difference between proportional and non-proportional quantifiers is reflected in the difference in the actual processing of quantifier expressions (see [23] and references given there). In these studies, it has been widely observed that proportional quantifiers take longer time to process and are interpreted less accurately than standard quantifiers. More specifically, Szymanik and Zajenkowski [23] argued that the computational identification of generalized quantifiers using semantic automata is relevant to cognitive verification processes of natural language quantifiers. In their experiment, participants were asked to judge whether quantified sentences were true of pictures containing 15 objects with different properties. Response times for the verification tasks were significantly longer in proportional quantified sentences (*more than half* and *less than half*) than in standard quantified sentences (*all* and *some*). This indicates that interpretation of proportional quantifiers requires more cognitive effort than interpretation of standard quantifiers, which is in accord with the computational model of generalized quantifiers.

In sum, proportional and non-proportional quantifiers show a different logical and cognitive behavior. The main question we are concerned with in this paper is whether such a difference between two kinds of quantification also appears in diagrammatic reasoning. To approach to this question, we will start, in the next section, with discussing how diagrams can represent sentences containing proportional quantifiers.

3 Diagrams of Proportional Quantifiers

3.1 A Problem in Diagrammatic Representation of *Most*

Euler diagrams represent sets of objects in terms of circles or closed curves and represent the inclusion and exclusion relations between sets by combining circles. Note here that circle sizes are irrelevant to the understanding of inclusion and

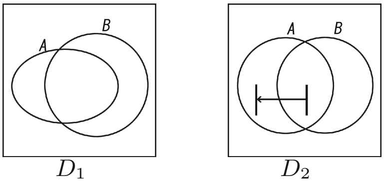


Fig. 1. A diagram for *Three-fourths of A are B* (D_1) and Rescher-Gallagher's Venn diagram with an arrow convention for *Most A are B* (D_2).

exclusion relations between sets i.e., the relations at the level (i) in our classification of set-theoretic relations discussed in Sect. 1. By contrast, the size differences play an important role in expressing the proportional relations between sets, i.e., the relations at the level (ii) in our classification.

As an example of a diagram in which the proportionality plays a role, consider the proportional quantifier *three-fourths* and its diagrammatic representation D_1 in Fig. 1. Here the area of the $A\overline{B}$ region is 1/4 of the total area of the A region and the area of the AB region is 3/4. Thus, this diagram corresponds to the sentence *Three-fourths of A are B*. We can find such a use of diagrams in the seminal work on conditionals by Adams [1], where the probabilities of conditionals are described by proportions of subregions in Euler-style diagrams. Also, in more recent developments of diagrammatic logic, the notion of "area-proportionality" is formalized using the weight values assigned to regions of Venn and Euler diagrams ([6]; see also [22]).

However, this line of extension of Euler diagrams is inadequate as a diagrammatic representation of the proportional quantifier *most*. In diagrammatic representations of *most*, one has to visualize the fact that the area of one region is greater than that of another region, without specifying the particular values of the areas of the relevant regions. This is an instance of the *over-specifity* property of diagrams discussed in details in [20,21]. What is necessary to capture the intended meaning of *most* is a natural device that indicates the proportionality relation between two regions and at the same time leaves underspecified the relation between the regions in question and the other regions in the diagram.

Rescher and Gallagher [15] overcame this problem by introducing to Venn diagrams the conventional device of the arrow to indicate that the extension of one region is less than that of another region. An example is shown as D_2 in Fig. 1. In this diagram, the AB region means the regions that can be extended, and the $A\overline{B}$ means the region that can be reduced. Thus even if we do not know the exact ratios of the regions' areas, we can extract the information that $\mathbf{A} \cap \mathbf{B}$ is greater than $\mathbf{A} - \mathbf{B}$.

By using the framework of Venn diagrams and introducing the conventional device of arrows, Rescher and Gallagher's diagrams succeed in avoiding the over-specificity problems inherent to proportional diagrams. However, giving up the idea of using the proportion of regions to indicate the proportionality of sets makes their diagrams less intuitive and hence more difficult to understand in

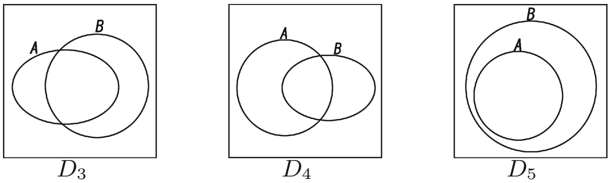


Fig. 2. Proportional diagrams for *Most A are B* (D_3), *Most A are not B* (D_4), and *All A are B* (D_5)

actual use. We can say that Venn diagrams with the arrow conventions are hybrid in that they combine a concrete circle-based form and an abstract convention in terms of arrows to represent the relational meaning of *most*.

For the diagrammatic representation of proportional quantifiers, we prefer to preserve the idea that the proportion of regions indicates the proportional relation between sets. In our view, diagrams that can be actually available to users are written on a paper or displayed on a PC monitor, with concrete forms. For this reason, in our experimental study, we do not adopt Venn diagrams with arrow conventions but instead use Euler diagrams whose regions have different areas, with the understanding that the sizes of the relevant regions whose proportionality is in focus are fixed and the other parts can be freely extended or reduced. We call the diagrams used in our experiments *proportional diagrams*.

In the actual scenes of the experiment, proportional diagrams were provided not as mere pictorial images, but rather as instances of general diagrams for *most* by instructing their syntax and semantics (see Appendix 2).[1] For example, *Most A are B* is represented by D_3 of Fig. 2, where (i) the proportion of AB region by $A\overline{B}$ region is specified as a *greater than* ($>$) relation with the tentative ratio of 2 : 1, and (ii) the proportion of BA region by $B\overline{A}$ region is unknown; the tentative ratio is set to 1 : 1, in order to restrain the invalid inference from *Most A are B* to *Most B are A*. D_4 of Fig. 2 corresponds to *More A are not B*, where the AB region and $A\overline{B}$ region are set to a 1 : 2 proportion and the BA region and $B\overline{A}$ region are set to a 1 : 1 proportion. D_5 of Fig. 2 corresponds to *All A are B*, where the area of the AB region is equal to that of the $\overline{A}B$ region and the invalid inference from *All A are B* to *Most B are A* is restrained.

Before moving on to discussing how proportional diagrams can be used in reasoning, two remarks are in order here. First, the system presented here works for pair-wise syllogistic inferences, where each sentence (premise and conclusion) has only two terms; however, without introducing a further convention, the system does not generalize to more complex cases. We made this simplifying assumption because we focused on the effectiveness of proportional diagrams in actual syllogistic inferences. To generalize the system of proportional diagrams, one would need a syntactic device to distinguish partially overlapping circles

[1] In addition, we can adopt a method using size-scalable diagrams in which object sizes can be changed from a default. Sato et al. [19] reported that the use of size-scalable diagrams in logical reasoning reduced the interfering effect of diagram layout.

expressing indeterminacy, as is common in Euler diagrams and Venn diagrams, from partially overlapping circles indicating the special proportionality relation expressing the meaning of *Most A are B*. One may use the Rescher-Gallagher's convention of using arrows or, for that matter, any syntactic convention to indicate the proportionality relation between two circles. Such a generalization of proportional diagrams and its empirical evaluation are left for future research.

Secondly, it is worth mentioning that Euler diagrams for proportional quantification used here can be naturally formalized within the framework of the relation-based approach to formalization of Euler diagrams [12,14]. According to this approach, a relation between sets such as inclusion and exclusion relations is taken as a primitive to define a diagram at an abstract level. Then a semantics and diagrammatic proof system can be provided in a similar way to the system of natural logic [13]. It is natural to add to this relation-based framework a relation corresponding to *Most A are B* (see [8] for a related study). We leave the detailed formal treatment of proportional diagrams for another occasion.

3.2 Reasoning with Proportional Diagrams

Reasoning with the proportional quantifier *most* has been studied in some depth by logicians having interest in natural language inferences (for early works, see [2]). Within syllogistic fragments, which are sometimes called as "plurative syllogisms", some researchers have proposed decision procedures with diagrams: Venn diagrams with arrow in [15], as stated above, and Lewis Carroll diagrams in [9]. More recently, the "natural logic" approach to natural language inference has combined natural language semantics based on the generalized quantifier theory, as seen in Sect. 2, and proof theory, to provide modern reconstructions of syllogistic reasoning. Endrullis and Moss [8] developed a proof system of syllogistic reasoning with *all*, *some*, and *most*.

Consider the following four arguments with *most*:

- *All BA, Most CB* (AU1); therefore, *Most CA* (U)
- *All AB, Most C not B* (AW2); therefore, *Most C not A* (W)
- *No BA, Most CB* (EU1); therefore, *Most C not A* (W)
- *No AB, Most CB* (EU2); therefore, *Most C not A* (W)

Here the label A stands for a sentence with *All*, E for a sentence with *No*, U for a sentence with *Most*, and W for a sentence with *Most-not*. The conclusion or hypothesis (beginning with *therefore*) is *entailed* by the premises in each argument. In other words, if all the premises are true, then the conclusion also is necessarily true. Therefore, each argument is a valid inference. Indeed, except for trivial cases (e.g., conclusion sentences converted from *All CA* to *Most CA*, and from *No CA* to *Most C not A*), valid syllogisms involving *most* and *most-not* comprise only the above four arguments (see [25]; for the sake of simplicity, *some* is not included here).

For invalid arguments, furthermore, non-entailment relations between premises and hypothesis can be generated in two ways (cf. [11], Chap. 5). First is a *contradiction*. Taking the AU1 syllogism above as an example, the hypothesis *Most C not A* contradicts the premises; if the premises are true, then the hypothesis cannot be true. Second is a *consistency* (*compatibility*). In the AU1 syllogism, the hypothesis *All CA* or *No CA* is consistent with the premises, in that if the premises are true, then the hypothesis may or may not be true. (See Appendix 1 for more details; note that not all valid/invalid syllogisms were used here.)

The extent to which ordinary people correctly make inferences with *most* in a typical sentential format is not well understood. If the interpretations of *most* are computationally and cognitively more complex than the interpretations of *all* (see Sect. 2), it is natural to speculate that inferences with *most* are also more complex than inferences with *all*. However, there is no empirical support for this speculation. It may be more accurate to say that there are no experimental studies which cover all of our tasks using *most*, *most-not*, *all*, and *no*. As a notable exception, Chater and Oaksford [5] employed the AU1U syllogism in their experiment, with a resulting accuracy rate of 85 %, and Geurts and van Der Slik [10] included inferences with *most* in more extended syllogisms with multiple quantifiers, for example, inferences from *Most A played against more than two B* and *All B were C* to *Most A played against more than two C*. Changes of *most* to *every* made no difference between them in participants' performances.

The solving process for reasoning tasks with proportional diagrams is essentially the same as that for standard syllogistic tasks with Euler diagrams [17]. Tasks for logical reasoning with diagrams typically consist of sentences (premises and a conclusion), and diagrams corresponding to the premise sentences, as shown in Figs. 3 and 4. The processes of unifying diagrams and extracting the information from sentences are illustrated by arrows.

Figure 3 shows the cases of an extended syllogism having the premises *All BA* and *Most CB*. In (1), (2) and (3), the first premise *All BA* is represented by D_1, and the second premise *Most CB* by D_2. There are two possible configurations of circles C and A in unifying the premise diagrams D_1 and D_2. In the first unified diagram D_3, the CA region is larger than the $C\overline{A}$ region. From this diagram we can extract the information $|\mathbf{C} \cap \mathbf{A}| > |\mathbf{C} - \mathbf{A}|$ (i.e., *Most CA*). In the second unified diagram D_4, circle C is totally included in circle A, thus we can extract the information $\mathbf{C} \subseteq \mathbf{A}$ (i.e., *All CA*).

Let us see how to solve the inferences in (1), (2) and (3) in turn. For (1), which is an instance of AU1U syllogism, one can start with extracting the information from the hypothesis *Most CA*. Then one can test whether this bottom-up information matches the top-down information extracted from the unified diagrams. Given that *All CA* implies *Most CA*, we can match the information from the hypothesis to both the information from D_3 and that from D_4. Thus, we can correctly judge that the hypothesis *Most CA* is entailed by the premises.

Regarding (2) (AU1W syllogism), the hypothesis has the form *Most C not A*. The information that can be read off from this hypothesis does not match the information from D_3 nor the information from D_4. Thus, we can judge that the hypothesis *Most C not A* contradicts the premises.

Regarding (3) (AU1A syllogism), the hypothesis has the form *All CA*. The information from the hypothesis matches information from D_4, but not from D_3 (*Most CA* of D_3 does not imply *All CA*)[2]. We can thus conclude that the given hypothesis *All CA* may be true or may be false. Thus, we can judge that the hypothesis *All CA* is consistent with the premises.

The above strategies that test whether the top-down information extracted from the unified diagrams matches the bottom-up information of the hypothesis sentence are common to standard syllogistic tasks using Euler diagrams. Figure 4 shows some examples. The syllogisms in (1) and (2) have the premises *All AB* and *No CB*. In this case, the unification of premise diagrams D_1 and D_2 produces the unique configuration D_3, corresponding to *No CA*. In the case of AE2E syllogism in (1), the hypothesis *No CA* matches the information *No CA* from the unified diagram D_3, leading to the correct answer (entailment). In the case of AE2A syllogism in (2), the hypothesis *All CA* does not match the information *No CA* from D_3. Accordingly, we can judge that the premises contradict the hypothesis. The syllogism in (3) (EA3E syllogism) has the premises *No BA* and *All BC* and the consistent conclusion *No CA*. There are four possibilities for the relationships between circles C and A: D_6, D_7, D_8, and D_9. The hypothesis matches D_6 but not D_7, D_8 or D_9. This kind of syllogism with no valid conclusion can actually be solved by enumerating multiple possibilities to unify the premise diagrams (see [18] for more details).

3.3 Predictions

Based on the analyses so far, we make two predictions.

(1) Reasoning with the proportional quantifier *most* (*most-not*) is more difficult and effortful than reasoning with standard quantifier *all* (*no*). This is true for participants who use only linguistic material as well as those who are also given proportional diagrams.

(2) The proportional diagrams improve the accuracy and speed of performances not only in reasoning with standard quantifiers but also in reasoning with proportional quantifiers, including all modes of entailment, contradiction, and consistency.

In this study, diagrams are added to sentential tasks of reasoning. This situation requires participants to handle both sentences and diagrams; i.e., they must do two jobs (cf. [16]). Nevertheless, if performance is faster with diagrams than without diagrams, this is evidence for the effectiveness of diagrams in reasoning. By contrast, even if performance is more accurate with than without diagrams, this does not necessarily count as evidence for the effectiveness of diagrams, because only the tasks with diagrams contain additional information on diagrams. Therefore, response-time data as well as accuracy data will be used to evaluate the effect of diagrams.

[2] Note here that the existence of $C\overline{A}$ region is a counter-example to the argument of the AU1A syllogism. See Takemura [24] for a formal specification of counter-example construction with Euler diagrams.

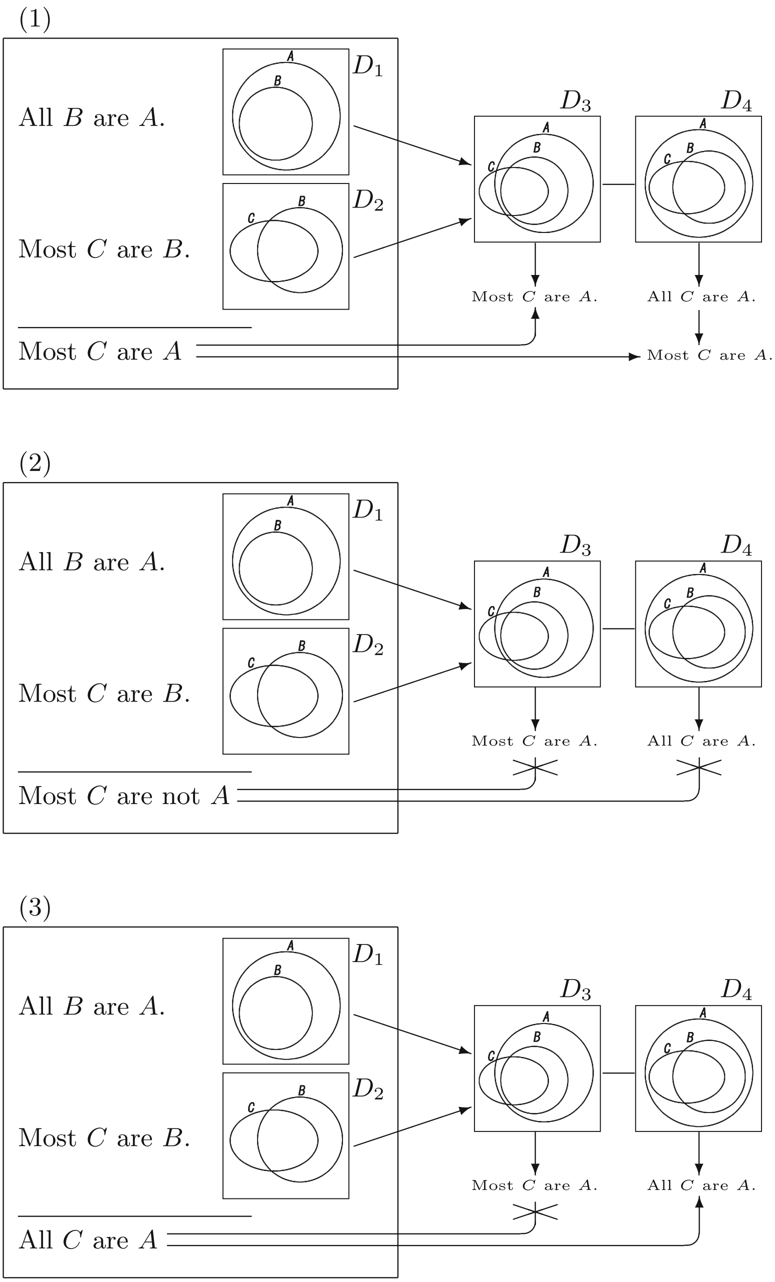


Fig. 3. Solving processes for reasoning tasks using proportional diagrams: (1) AU1U syllogism: entailment, (2) AU1W syllogism: contradiction, (3) AU1A syllogism: consistency

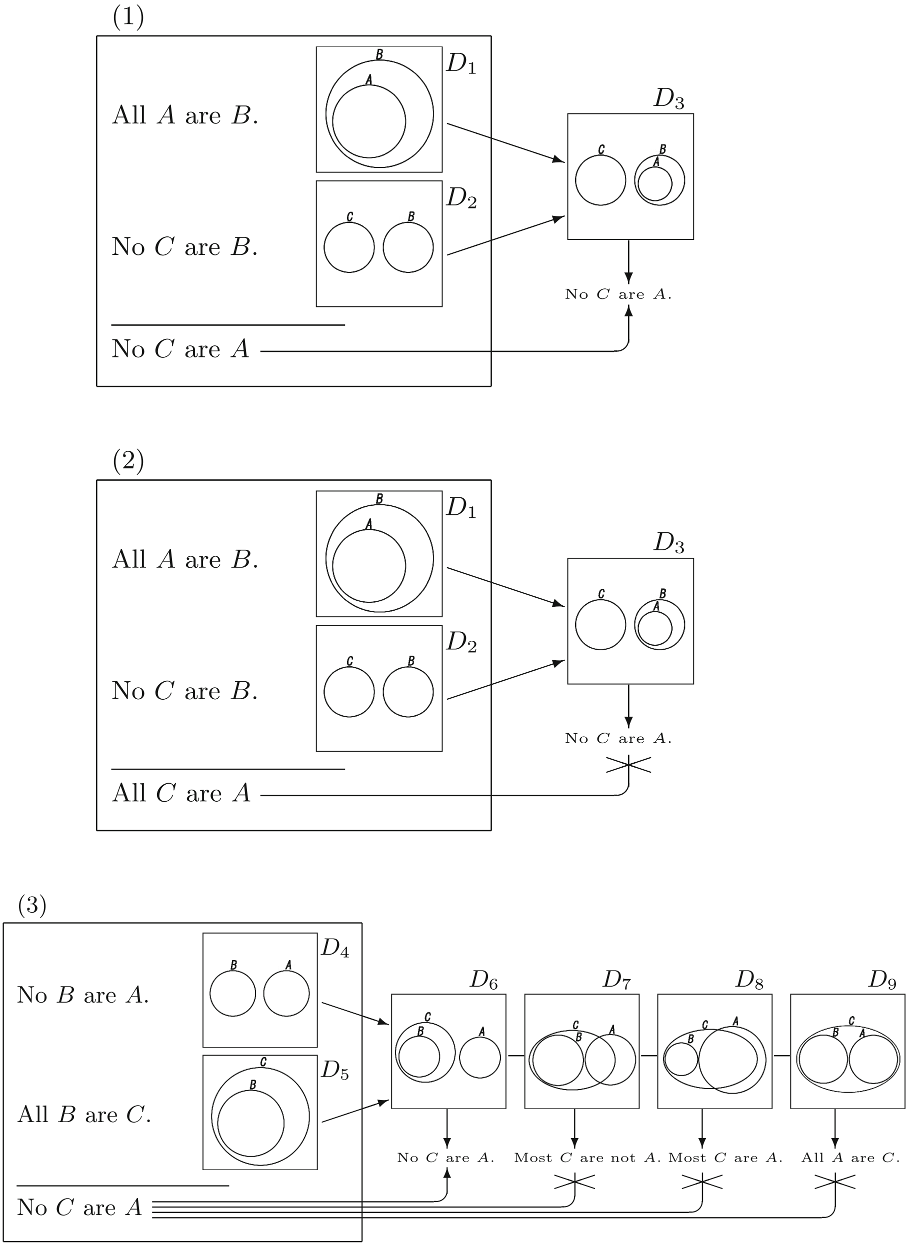


Fig. 4. Solving processes for reasoning tasks using Euler diagrams: (1) AE2E syllogism: entailment, (2) AE2A syllogism: contradiction, (3) EA3E syllogism: consistency

4 Experiment

4.1 Method

Participants. Forty-two undergraduate and graduate students from the University of Tokyo were recruited by means of an advertisement posted on campus. The mean age was 20.33 ($SD = 2.09$) with a range of 18–28 years. All participants gave informed consent and were paid for their participation. The Ethics Review Committee of the Graduate School of Arts and Sciences at the University of Tokyo approved all procedures in this experiment. The participants were Japanese-speaking students, and the sentences and instructions were provided in Japanese. None had any prior training in syllogistic logic. Participants were divided into two groups: a Diagrammatic group ($N = 21$), in which diagrams were used, and a Linguistic group ($N = 21$), in which diagrams were not used.

Materials. We presented 39 items: 17 standard syllogisms and 22 non-standard syllogisms (see Appendix 1 for the list of syllogisms). The sentences of standard syllogisms were universally quantified sentences either of the form *All A are B* or *No A are B*. The sentences of non-standard syllogisms were proportional quantified sentences of the forms *Most A are B* and *Most A are not B*. As shown in Fig. 5, the participants were presented with two premises and a hypothesis (conclusion) on a PC monitor and were asked to answer the question of *If the following two premises are true, is the hypothesis also true?*, by selecting a response from a list of three options: 1. *Hypothesis is true* (i.e., entailment). 2. *Hypothesis is false* (contradiction). 3. *Neither 1 nor 2: Hypothesis may or may not be true* (consistency). The premises in 9 syllogisms (4 non-standard) entail the hypotheses and the premises in 13 syllogisms (7 non-standard) contradict the hypotheses, and the premises in 17 syllogisms (11 non-standard) are consistent with the hypotheses. The quantified sentences included three properties, color (red or blue) for A terms, shape (square or round) for B terms, and striped pattern (horizontal or vertical) for C terms.

Procedures. The experiment was conducted individually. First, the Diagrammatic group only was given two pages of instructions on the meaning of Euler diagrams, but they did not receive any instructions about how to manipulate diagrams when solving syllogisms (for details, see Appendix 2). Second, both groups were given two pages of instructions regarding on three types of entailment relationships between premises and conclusion: “entailment”, “contradiction” and “consistency” each with an example (for details, see Appendix 2). The participants were asked to press, as quickly and accurately as possible, a button with the number representing their answer. The 39 reasoning tasks were presented in random order. There was no time limit.

4.2 Results and Discussion

Accuracy rates (numbers of correct answers) for the non-standard syllogisms were significantly lower than those for the standard syllogisms. This tendency

If the following two premises are true,
is the hypothesis also true?

Premise 1 All round objects are blue.

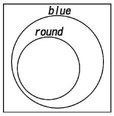
Premise 2 Most vertical objects are round.

Hypothesis Most vertical objects are blue.

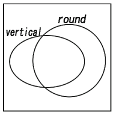
1. Hypothesis is true.
2. Hypothesis is false.
3. Neither 1 nor 2: Hypothesis may or may not be true.

If the following two premises are true,
is the hypothesis also true?

Premise 1 All round objects are blue.



Premise 2 Most vertical objects are round.



Hypothesis Most vertical objects are blue.

1. Hypothesis is true.
2. Hypothesis is false.
3. Neither 1 nor 2: Hypothesis may or may not be true.

Fig. 5. Examples of reasoning tasks (AU1U syllogism) for the Linguistic group (left) and the Diagrammatic group (right)

was common to both the Linguistic group (73.4 % vs. 84.9 %, $z = 2.78$, $P = 0.0055$, Wilcoxon test) and the Diagrammatic group (81.6 % vs. 92.4 %, $z = 2.80$, $P = 0.0051$). Response times (for correctly answered items) for the non-standard syllogisms were also significantly longer than those for the standard syllogisms. This tendency was also common to both the Linguistic group (26.98 s vs. 21.05 s, $t(20) = 5.074$, $p < 0.01$) and the Diagrammatic group (20.13 s vs. 14.19 s, $t(20) = 6.084$, $p < 0.01$). Thus, syllogistic reasoning with proportional quantifiers *most* is relatively difficult and effortful, supporting our first prediction.

Table 1 shows accuracy rates and response times (for correct answers only) in the Linguistic and Diagrammatic groups. The results for each syllogistic type are shown in Appendix 1. For the non-standard syllogisms, accuracy rates (numbers of correct answers) were substantially higher in the Diagrammatic group than in the Linguistic group (81.6 % vs. 73.4 %), but there was no significant difference between them ($U = 146.5$, $P = 0.0610$, Mann-Whitney Test). Accuracy rates for the standard syllogisms were significantly higher in the Diagrammatic group than in the Linguistic group (92.4 % vs. 84.7 %, $U = 146.5$, $P = 0.0422$). More detailed analyses of entailment, contradiction, and consistency were conducted. For the non-standard syllogisms, accuracy rates were significantly higher in the Diagrammatic group than in the Linguistic group for entailment (90.5 % vs. 80.9 %, $U = 142.5$, $P = 0.0233$) and contradiction (88.4 % vs. 80.3 %, $U = 137$, $P = 0.0275$). There was no significant difference for consistency (74.0 % vs. 66.2 %, $U = 177$, $P = 0.2669$). For the standard syllogisms, there was no significant difference for each condition of entailment ($P = 0.2681$), contradiction ($P = 0.3425$) and consistency ($P = 0.0586$).

Response times of correctly answered items were logarithmically transformed and subjected to t-tests. Response times were significantly shorter in the Diagrammatic group than in the Linguistic group for standard syllogisms (14.19 s vs. 21.05 s, $t(40) = 3.506$, $p < 0.01$) and for non-standard syllogisms (20.13 s vs.

Table 1. Accuracy rates and response times (correct answer only) for standard syllogisms and non-standard syllogisms in the Linguistic group and Diagrammatic group

	Entailment	Contradiction	Consistency	Total
Standard syllogisms				
Linguistic group	92.4 % 20.38 s	92.1 % 18.37 s	71.4 % 24.79 s	84.9 % 21.05 s
Diagrammatic group	96.2 % 12.74 s	95.2 % 13.77 s	86.5 % 15.55 s	92.4 % 14.19 s
Non-standard syllogisms				
Linguistic group	80.9 % 22.16 s	80.3 % 23.48 s	66.2 % 31.90 s	73.4 % 26.98 s
Diagrammatic group	90.5 % 16.68 s	88.4 % 17.96 s	74.0 % 23.89 s	81.6 % 20.13 s

26.98 s, $t(40) = 2.526$, $p < 0.05$). In the following analyses, we excluded the participants' data where there was no correct answer in each condition of entailment, contradiction, and consistency. In the standard syllogisms, response times were significantly shorter in the Diagrammatic group than in the Linguistic group for entailment (12.74 s vs. 20.38 s, $t(39) = 5.566$, $p < 0.01$), contradiction (13.77 s vs. 18.37 s, $t(39) = 2.205$, $p < 0.05$), and consistency (15.55 s vs. 24.79 s, $t(39) = 3.887$, $p < 0.01$). In the non-standard syllogisms, response times were significantly shorter in the Diagrammatic group than in the Linguistic group for entailment (16.68 s vs. 22.16 s, $t(39) = 2.095$, $p < 0.05$), contradiction (17.96 s vs. 23.48 s, $t(39) = 2.393$, $p < 0.05$), and consistency (23.89 s vs. 31.90 s, $t(39) = 2.127$, $p < 0.05$). The results shown in Table 1 suggest that proportional diagrams tend to be effective in syllogistic reasoning with the proportional quantifier *most*, consistent with our second prediction.

5 Conclusion and Future Work

As seen in Sect. 2, previous work has revealed that *most* sentences are computationally and cognitively complex in interpretation or verification. Little attention, however, has been paid to *most* sentence inferences; further, not much is yet known about the kind of diagrams that work well in such inferences. In this study, we showed that syllogistic reasoning with proportional quantifiers *most* is cognitively complex yet can effectively be supported by Euler-style diagrams that represent the proportionality relationships between sets in terms of area-proportionality. In particular, our result indicates that difference between *all* and *most* in the complexity of comprehension also reflects the complexity of reasoning tasks in both linguistic and diagrammatic formats.

Future research should explore the interaction of proportional diagrams and diagrams for existential quantifiers. For instance, it is well-discussed that *Most A are B* and *Most A are C* entail *Some B are C* [9]. It is clear that diagrams expressing the proportionality relationship in an intuitive way can help deriving such an inference from *most* to *some*. However, combining proportional diagrams with diagrams asserting the non-emptiness of a set is not a trivial task; in addition to the generalization of proportional diagrams mentioned in Sect. 3.1, we leave for future research how to set up a more expressive representation system for proportional diagrams.

In the study of visualization and graphics, it has been discussed that perceptual judgements of the relative sizes of areas is relatively difficult and effortful [7]. A detailed investigation on perception of proportional diagrams is also left for future work.

As we saw in Sect. 2, by merging the theoretical and empirical findings on various kinds of quantifiers, we can explore the relationship between logical, computational and cognitive aspects of human reasoning. Imposing a constraint on the possible ways in which we can reason by means of diagrams can contribute to this direction of research and serve as a fruitful way to capture the complexity of reasoning tasks.

Appendix 1: The Results of Each Task

Table 2. Accuracy rates and response times (correct answer only) for 39 syllogisms in the Linguistic group (left) and Diagrammatic group (right)

Premises	Entailment	Contradiction	Consistency
All BA, All CB (AA1)	All CA (A) 100 % 100 % 15.59 s 14.68 s	No CA (E) 100 % 100 % 14.75s 13.18 s	-
All AB, All CB (AA2)	-	-	No CA (E) 52.4 % 90.5 % 25.93 s 17.87 s
All BA, All BC (AA3)	-	No CA (E) 85.7 % 95.2 % 17.34 s 17.95 s	All CA (A) 80.9 % 90.5 % 16.88 s 30.10s
All BA, No CB (AE1)	-	-	No CA (E) 71.4 % 90.5 % 25.23 s 32.55 s
All AB, No CB (AE2)	No CA (E) 85.7 % 95.2 % 21.93 s 22.55 s	All CA (A) 85.7 % 95.2 % 23.59 s 21.73 s	-
All BA, No BC (AE3)	-	-	no CA (E) 95.2 % 85.7 % 23.95 s 12.28 s
All AB, No BC (AE4)	No CA (E) 81.0 % 90.5 % 29.46 s 27.06 s	All CA (A) 90.5 % 90.5 % 25.11 s 15.94 s	-
No BA, All CB (EA1)	No CA (E) 100 % 100 % 16.48 s 13.88 s	All CA (A) 100 % 100 % 15.53 s 10.72 s	-
No AB, All CB (EA2)	No CA (E) 95.2 % 95.2 % 21.29 s 9.77 s	All CA (A) 90.5 % 90.5 % 19.50 s 15.92 s	-
No BA, All BC (EA3)	-	-	No CA (E) 66.7 % 80.9 % 27.83 s 25.22 s
No AB, All BC (EA4)	-	-	No CA (E) 61.9 % 80.9 % 30.28 s 19.67 s
All BA, Most CB (AU1)	Most CA (U) 90.5 % 95.2 % 17.85 s 22.79 s	Most C not A (W) 90.5 % 95.2 % 16.48 s 15.89 s	All CA (A) 57.1 % 76.2 % 22.23 s 19.52s
All AB, Most CB (AU2)	-	All CA (A) 61.9 % 66.7 % 24.78 s 31.06 s	Most CA (U) 85.7 % 90.5 % 26.56 s 24.49 s
All BA, Most BC (AU3)	-	-	Most C not A (W) 61.9 % 66.7 % 23.16 s 35.39 s
All BA, Most C not B (AW1)	-	No CA (E) 61.9 % 90.5 % 26.29 s 38.56 s	Most C not A (W) 71.4 % 76.2 % 24.87 s 22.92 s
All AB, Most C not B (AW2)	Most C not A (W) 61.9 % 80.9 % 24.11 s 19.12 s	Most CA (U) 90.5 % 85.7 % 25.70 s 12.87 s	No CA (E) 52.4 % 80.9 % 27.26 s 33.68 s
All BA, Most B not C (AW3)	-	-	Most CA (U) 76.2 % 76.2 % 36.51 s 20.29 s
All AB, Most B not C (AW4)	-	-	Most C not A (W) 71.4 % 71.4 % 50.49 s 23.45 s
No BA, Most CB (EU1)	Most C not A (W) 100 % 90.5 % 25.72 s 17.67 s	Most CA (U) 95.2 % 90.5 % 19.69 s 20.26 s	No CA (E) 52.4 % 71.4 % 19.42 s 13.37 s
No AB, Most CB (EU2)	Most C not A (W) 71.4 % 95.2 % 21.46 s 39.25 s	Most CA (U) 90.5 % 95.2 % 20.89 s 16.95 s	No CA (E) 57.1 % 71.4 % 35.34 s 30.15 s
No BA, Most BC (EU3)	-	-	Most C not A (W) 81.0 % 61.9 % 27.60 s 15.60 s
No AB, Most BC (EU4)	-	All CA (A) 71.4 % 95.2 % 36.81 s 18.59 s	Most CA (U) 61.9 % 71.4 % 31.32 s 35.92 s

Appendix 2: Instructions Used in Experiment

See: http://abelard.flet.keio.ac.jp/person/sato/index/appendix_d16.pdf

References

1. Adams, E.W.: The Logic of Conditionals: An Application of Probability to Deductive Logic. Springer, Dordrecht (1975)
2. Altham, J.E.J.: The Logic of Plurality. Methuen, London (1971)
3. Barwise, J., Cooper, R.: Generalized quantifiers and natural language. Linguist. Philos. **4**, 159–219 (1981)
4. van Benthem, J.: Essays in Logical Semantics. Reidel, Dordrecht (1986)
5. Chater, N., Oaksford, M.: The probability heuristics model of syllogistic reasoning. Cogn. Psychol. **38**, 191–258 (1999)
6. Chow, S., Ruskey, F.: Drawing area-proportional Venn and Euler diagrams. In: Liotta, G. (ed.) GD 2003. LNCS, vol. 2912, pp. 466–477. Springer, Heidelberg (2004)
7. Cleveland, W.S., McGill, R.: Graphical perception: theory, experimentation, and application to the development of graphical methods. J. Am. Stat. Assoc. **79**, 531–554 (1984)
8. Endrullis, J., Moss, L.S.: Syllogistic logic with "most". In: de Paiva, V., de Queiroz, R., Moss, L.S., Leivant, D., de Oliveira, A. (eds.) WoLLIC 2015. LNCS, vol. 9160, pp. 124–139. Springer, Heidelberg (2015)
9. Geach, P.T.: Reason and Argument. University of California Press, Berkeley (1976)
10. Geurts, B., van Der Slik, F.: Monotonicity and processing load. J. Seman. **22**, 97–117 (2005)
11. MacCartney, B.: Natural Language Inference. Ph.D. thesis, Stanford University (2009)
12. Mineshima, K., Okada, M., Takemura, R.: A diagrammatic reasoning system with Euler circles. J. Logic Lang. Inf. **21**, 365–391 (2012)
13. Mineshima, K., Okada, M., Takemura, R.: A generalized syllogistic inference system based on inclusion and exclusion relations. Stud. Logica **100**, 753–785 (2012)
14. Mineshima, K., Sato, Y., Takemura, R., Okada, M.: Towards explaining the cognitive efficacy of Euler diagrams in syllogistic reasoning: a relational perspective. J. Vis. Lang. Comput. **25**, 156–169 (2014)
15. Rescher, N., Gallagher, N.A.: Venn diagrams for plurative syllogisms. Philos. Stud. **16**, 49–55 (1965)
16. Sato, Y., Masuda, S., Someya, Y., Tsujii, T., Watanabe, S.: An fMRI analysis of the efficacy of Euler diagrams in logical reasoning. In: VL/HCC 2015, pp. 143–151. IEEE Press (2015)
17. Sato, Y., Mineshima, K.: How diagrams can support syllogistic reasoning: an experimental study. J. Logic Lang. Inf. **24**, 409–455 (2015)
18. Sato, Y., Wajima, Y., Ueda, K.: An empirical study of diagrammatic inference process by recording the moving operation of diagrams. In: Dwyer, T., Purchase, H., Delaney, A. (eds.) Diagrams 2014. LNCS, vol. 8578, pp. 190–197. Springer, Heidelberg (2014)
19. Sato, Y., Wajima, Y., Ueda, K.: Visual bias of diagram in logical reasoning. In: CogSci 2014, pp. 2342–2347. Cognitive Science Society, Austin (2014b)

20. Shimojima, A.: On the Efficacy of Representation. Ph.D. thesis, Indiana University (1996)
21. Shimojima, A.: Semantic Properties of Diagrams and Their Cognitive Potentials. CSLI Publications, Stanford (2015)
22. Stapleton, G., Rodgers, P., Howse, J.: A general method for drawing area-proportional Euler diagrams. J. Vis. Lang. Comput. **22**, 426–442 (2011)
23. Szymanik, J., Zajenkowski, M.: Comprehension of simple quantifiers: empirical evaluation of a computational model. Cogn. Sci. **34**, 521–532 (2010)
24. Takemura, R.: Counter-example construction with Euler diagrams. Stud. Logica **103**, 669–696 (2015)
25. Thompson, B.: Syllogisms using “few”, “many”, and “most”. Notre Dame J. Form. Logic **23**, 75–84 (1982)