J Log Lang Inf (2015) 24:409–455
DOI 10.1007/s10849-015-9225-4

How Diagrams Can Support Syllogistic Reasoning: An Experimental Study

Yuri Sato[1] · **Koji Mineshima**[2]

Published online: 8 October 2015

Abstract This paper explores the question of what makes diagrammatic representations effective for human logical reasoning, focusing on how Euler diagrams support syllogistic reasoning. It is widely held that diagrammatic representations aid intuitive understanding of logical reasoning. In the psychological literature, however, it is still controversial whether and how Euler diagrams can aid untrained people to successfully conduct logical reasoning such as set-theoretic and syllogistic reasoning. To challenge the negative view, we build on the findings of modern diagrammatic logic and introduce an Euler-style diagrammatic representation system that is designed to avoid problems inherent to a traditional version of Euler diagrams. It is hypothesized that Euler diagrams are effective not only in interpreting sentential premises but also in reasoning about semantic structures implicit in given sentences. To test the hypothesis, we compared Euler diagrams with other types of diagrams having different syntactic or semantic properties. Experiment compared the difference in performance between syllogistic reasoning with Euler diagrams and Venn diagrams. Additional analysis examined the case of a linear variant of Euler diagrams, in which set-relationships are represented by one-dimensional lines. The experimental results provide evidence supporting our hypothesis. It is argued that the efficacy of diagrams in supporting syllogistic reasoning crucially depends on the way they represent the relational information contained in categorical sentences.

✉ Yuri Sato
satoyuri0@gmail.com

Koji Mineshima
mineshima.koji@ocha.ac.jp

1 Department of General Systems Studies, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan

2 Center for Simulation Sciences, Ochanomizu University, Tokyo, Japan

Keywords Efficacy of diagrams · Human reasoning · Logic and cognition · Human experimentation · Quantification · Categorical syllogism · External representation

1 Introduction

Humans have a familiarity with various types of diagrammatic representation in reasoning and problem solving. It is relatively unproblematic that diagrammatic representations work well for reasoning about concrete objects located in space. For example, various kinds of graphical representations such as maps are effective tools to grasp one's present position in a city or to search for the shortest route to a destination (Meilinger and Knauff 2008; Gattis 2005). Even in an elementary physics problem such as a pulley problem, it would be difficult to understand how objects move if the relevant illustrations are not presented (Cheng 2004; Heiser and Tversky 2006; Hegarty 2004). Over the past few decades, the effectiveness of diagrams in such spatial reasoning tasks has been the subject of research interest in cognitive studies (Glasgow et al. 1995; Purchase 2014). One crucial observation emphasized by various authors (e.g., Palmer 1978; Myers and Konolige 1994; Barwise and Hammer 1996; Gurr et al. 1998) is that the effectiveness of diagrammatic representations depends on a certain kind of similarity or correspondence relation holding between the representations and what they represent, for example, a relation between maps and concrete objects having spatial properties. In the case of reasoning with concrete objects located in space, diagrams and their target domains are both spatial objects; hence, although there is a debate on the nature of its underlying formal semantics (cf. Pratt 1993; Camp 2007; Rescorla 2009), it is relatively clear that a certain kind of correspondence relation can hold between a spatial diagram like a map and what it represents. If such a relation holds, the reasoning required to solve a given task could be visually simulated in terms of concrete operations on diagrams, resulting in the effective use of diagrams in reasoning. However, it is not clear that the same account can be applied to cases of more general and abstract reasoning. Arguably, the less specific a graphical representation becomes, the less effective it is to simulate a reasoning task in question (cf. Stenning and Oberlander 1995). Logical reasoning is a paradigmatic example of higher-level reasoning that has attracted the interest in cognitive science. A question then arises: how can logical reasoning be supported by diagrammatic representations?

What is puzzling here is the fact that logic is by its nature topic-neutral and general-purpose; it must be applicable to arbitrary domains, ranging from empirical domains such as physics and chemistry to more abstract mathematical domains (for a discussion on the topic-neutrality or generality of logic, see Hodes 1984; MacFarlane 2002). The dilemma is that while logical reasoning is quite general in that there is no particular domain to which it is dedicated, effective diagrams must be concrete objects that can be manipulated by human in an intuitive way. More specifically, the validity of a logical inference is determined not by the meaning of content words such as *cat* or *human* but by the meaning of logical expressions such as connectives (most paradigmatically, *and*, *or*, *if*, and *not*) and quantifiers (*all*, *some*, *most*, and so on). Thus, to obtain an effective representation for logical reasoning, one has to visualize the very meaning of logical connectives and quantifiers, whether they are combined with concrete terms as in *All artists are beekeepers* or with schematic variables as in *All A are B*. It seems not clear at all what kind of a similarity or correspondence relation could hold between concrete diagrams and such abstract meanings of logical expressions.

To understand the effectiveness of diagrams in logical reasoning, then, it will be necessary to elucidate the meaning of a statement used in logical reasoning and examine how it could be systematically represented by suitable diagrams. Throughout this paper, we will focus on the case of syllogistic reasoning, a paradigmatic example of logical reasoning that have attracted interests in cognitive science, and consider how it could be effectively supported by various forms of external diagrams. There is a well-known family of diagrammatic representations, namely, the so-called Euler diagrams (Euler 1768; Hammer and Shin 1998), which have been widely used as a tool for teaching students syllogistic and set-theoretical reasoning. Despite their acceptance, it still remains to be answered why Euler diagram are particularly effective in supporting logical reasoning. Indeed, it has been reported in the psychology literature that some forms of Euler diagrams do not improve but sometimes retard a certain class of syllogistic inferences (Rizzo and Palmonari 2005; Calvillo et al. 2006).[1] Thus it is fair to say that it is still controversial whether Euler diagrams are actually effective for untrained people to conduct syllogistic reasoning in a successful way.

Another well-known diagrammatic representation systems for logical reasoning is the so-called Venn diagrams (Venn 1881; Edwards 2004). As compared with Euler diagrams, Venn diagrams are syntactically more involved but have a semantically more transparent system (Shin 1994; see also Sect. 2.2 below). Indeed, it has been reported that Venn diagrams are actually effective in such tasks as searching and recognition (Michard 1982; Hertzum and Frøkjær 1996; Jones et al. 1999). In the context of cognitive psychology, however, less attention has been paid to Venn diagrams than to Euler diagrams. Specifically, there seem to be very few psychological experiments to compare how reasoning with Euler and Venn diagrams could contribute to solving logical reasoning tasks. Accordingly it still remains unclear how the difference in the design of diagrammatic syntax and semantics affects the cognitive efficacy of diagrammatic reasoning systems.[2] We will show that Euler diagrams have some advantages over Venn diagrams in solving syllogistic reasoning tasks; such a comparison would contribute to specifying the condition under which diagrams can be effective in logical reasoning.

[1] Bauer and Johnson-Laird (personal communication) in their unpublished study also reported that while Euler diagrams—or, more specifically, what we will classify as a traditional form of Euler diagrams (Euler circles)—improved reasoning with a class of difficult syllogisms but retarded reasoning with a class of easy syllogisms.

[2] It is standard in the literature of diagrammatic logic to distinguish between two layers of diagrammatic syntax: concrete (token) syntax and abstract (type) syntax (Howse et al. 2002). The former is concerned with the surface structure of a diagram that is visible to users, while the latter gives a formal definition that concrete diagrams must obey. Since we are concerned with cognitive properties of concrete diagrams presented to users, by (diagrammatic) syntax we mean concrete syntax throughout this paper. This is consistent with Stenning and Lemon's (2001) cognitive view. They claimed that a concatenation is self-evident in a diagrammatic representation, and thus semantic information can be *directly* interpreted from a diagrammatic representation without the mediation of abstract syntax.

To solve a logical reasoning task involves two kinds of process: one has to interpret sentential premises, and, based on the information obtained thereby, one has to draw a valid conclusion. Thus, to understand how diagrams aid logical reasoning, it is of central importance to give a suitable semantical analysis to linguistic materials composing a logical reasoning task in question. Focusing on the case of syllogistic reasoning, we will argue that elementary quantificational sentences composing syllogistic inferences can best be analyzed as expressing *relational* semantic information, in particular, inclusion and exclusion relations concerning sets and individuals; more details will be described in Sect. 2.3.1 (see also Mineshima et al. 2014 for a formal background). We claim that the key role played by diagrams in supporting syllogistic reasoning is to make explicit what binary relations are delivered by quantified sentences in a form that is suitable for manipulation. This claim is consistent with the view that reasoning with relational information plays an important role in various domains related to human higher cognition (cf. Halford et al. 2010). Cognitive properties and complexities of relational thinking have been studied in depth in the literature (Halford et al. 1998; Doumas et al. 2008). It is well known that quantifiers in natural languages denote a binary (higher-order) relation between sets (Barwise and Cooper 1981); but the well-observed fact that logically untrained reasoners tend to find a difficulty in solving a certain class of categorical syllogisms suggests that the relational information in question is not necessarily transparent to the reasoners (Khemlani and Johnson-Laird 2012). Our claim is that the effectiveness of Euler diagrams rests in the fact that they represent relational information about sets and individuals in a way that can support syllogistic reasoning. If our hypothesis is correct, it would be expected that any diagram that makes explicit the relational information of a quantified sentence in a suitable way would be effective in supporting syllogistic reasoning. Specifically, we predict that not only a diagram composed of two-dimensional objects such as circles but also one-dimensional version of diagrams (linear diagrams) could support syllogistic reasoning in equally effective ways. The study of linear diagrams will provide further empirical support to our theoretical claim.

The use of Euler diagrams in logical reasoning has attracted formal research attention since 1990s (Barwise and Etchemendy 1996), and the relevant various systems have been proposed and studied using the method of mathematical logic (e.g., Shin 1994; Howse et al. 2005). Currently, however, the cognitive underpinning of such systems has been seldom investigated experimentally; thus little is known about the cognitive efficacy of Euler and Venn diagrams in syllogistic reasoning. On the other hand, Euler diagrams have been also attracted the attention in the context of experimental psychology of syllogistic reasoning. Most notably, Erickson (1974) proposed that people use a mental analogue of Euler diagrams in solving syllogistic reasoning tasks. In such studies, however, Euler diagrams were regarded as a tool for analyzing the performance of syllogistic reasoning, not as an external device to support human reasoning. Furthermore, a traditional system of Euler diagrams, which is known in the psychological literature, have some inherent logical problems as we will discuss in Sect. 2.2.1. In our view, it is important to take into account the formal findings of the recent development of diagrammatic logic as mentioned above and then to test empirically how and whether (a modern version of) Euler diagrams can support syllogistic reasoning with quantified sentences.

The structure of this paper is as follows. First, Sect. 2 provides a theoretical framework of the efficacy of diagrams in logical reasoning. Based on the semantical and syntactic analysis of the Euler diagrammatic representations, we conduct a process analysis of syllogistic sentences interpretation and syllogistic inference with the diagrams. Sections 3 and 4 present an empirical study of efficacy of Euler diagrams. Section 3 describes an experiment comparing participants' performance in syllogism solving using Euler diagrams and Venn diagrams. Section 4 examines whether the pattern of results in Sect. 3 also emerges when a linear variant of Euler diagrams is used. Section 5 discusses the conclusions that can be drawn from the current findings.

2 Cognitive Model of Logical Reasoning with Diagrams

2.1 Two Kinds of Efficacy

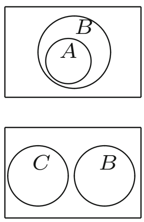
A task of logical reasoning, most typically, a task of checking the validity of an inference, involves multiple processes. Among them, we focus on two processes, namely, interpretation and inference, based on a standard two-stage framework in natural language semantics (see, e.g. Blackburn and Bos 2005; see also Stenning and van Lambalgen 2001, 2004, where the distinction of two kind of reasoning was emphasized: reasoning *toward* an interpretation of premises versus reasoning *from* a fixed interpretation of premisses). The reasoner first interprets the premise sentences, and thereby semantic information is obtained from sentences. Then the reasoner checks the validity of the argument using inferential mechanisms operating the semantic information in the appropriate way. In fact, it is well known that the reasoner actually makes errors in both the processes of interpretation and inference (cf. Evans et al. 1993). When solving logical problems with external diagrams, we assume that certain diagrams work well in both the two processes. A typical example of logical reasoning task using external diagrams is shown in Fig. 1. Here a syllogism is presented with Euler diagrams. How can such diagrams contribute to checking the validity of a logical argument? As stated in the above case of sentential reasoning, diagrams are also associated with semantic information, but in this case constitute syntactic objects to be manipulated in inferential processes. In this situation, two ways in which diagrams can be effective in logical reasoning can be highlighted.

Firstly, diagrams can help to improve interpretations of sentences, thereby avoiding logical reasoning errors due to misinterpretation. We call this kind of efficacy *interpre-*

Fig. 1 An example of syllogistic reasoning task with Euler diagrams

All A are B.

No C are B.



(Therefore, No C are A)

tational efficacy. For example, the sentence “All A are B” is commonly misinterpreted as equivalent to “All B are A”, an error known as the *illicit conversion error* in the literature (e.g. Chapman and Chapman 1959). However, participants presented with diagrams like those in Fig. 1 can immediately see that the diagrams corresponding to these two sentences are topologically different, and hence deliver different semantic information (cf. the discussion of (a)symmetricity of diagrams in Stenning 2002). In our view, such processes are formulated as processes of matching the semantic information obtained from diagrams with that obtained from sentences. In this case, the validity of an argument is checked based on the same type of process used in linguistic reasoning. Here diagrams are used in a static way, merely as a record of information (Barwise and Etchemendy 1996).

Secondly, and more importantly, diagrams can play a crucial role in inference processes themselves. We call this kind of efficacy *inferential efficacy*. For example, the process of solving logical reasoning tasks can be replaced by manipulation of diagrams, constructing “diagrammatic inference” to check the validity of a logical argument. We assume that these constructions are conducted through a proof-theoretical component of diagrammatic reasoning. If manipulating diagrams consists of simple and intuitive steps, this method may be more tractable than usual linguistic inference.

Indeed, Shimojima and Katagiri (2013) argue for the existence of “inference by hypothetical drawing” involving imaginary drawing of the given diagrams, based on eye-tracking data of subjects working with position diagrams in transitive inferential tasks. The inferential efficacy is consistent with the influential view in the study of external representations in general, that external representations can change the nature of tasks, namely, tasks with and without such external representations are completely different from users’ point of view (see Zhang and Norman 1994; Scaife and Rogers 1996). On the contrary, a negative view of the efficacy of diagrams in inferential processes was also put forward in a seminal study by Larkin and Simon (1987), which analyzed the three processes of *search*, *recognition*, and *inference* in diagrammatic reasoning. They claim that inference is largely independent of the way in which information is represented, and diagrams are less beneficial in inference. As such, it may be important to deal with the question whether or not there is the inferential efficacy.

Although our focus is on interpretational and inferential aspects of syllogistic reasoning supported by diagrams, there are other possible factors contributing to effective use of diagrams in logical reasoning. First, one has to *construct* diagrams that are suitable for a given reasoning problem, unless such diagrams are externally presented in advance. In the case of syllogistic reasoning, it has been reported in protocol studies of syllogistic reasoning (Ford 1994; Bucciarelli and Johnson-Laird 1999) that some of the participants tended to spontaneously draw circles to represent the relationship between sets. Since we are concerned with tasks where diagrams are externally provided (cf. Fig. 1), issues surrounding self-construction process of diagrams will be left aside in this paper. Second, as Gurr et al. (1998) emphasized, in addition to a process of constructing the desired diagram for a given reasoning task, a process of *recognizing* and *extracting* the relevant information in the constructed diagram plays a role in diagrammatic reasoning. Various conditions that are potentially relevant to the ease of information extraction from diagrams include wellformedness properties (Fish et al. 2011) and visual layouts (Benoy and Rodgers 2007); see Sato et al. (2011) for an empirical study of information extraction processes for Euler and Venn diagrams.

In Sect. 2.2, we will introduce diagrammatic representation systems (Euler and Venn diagrams) used in our experiments. Then, in Sect. 2.3, we will discuss how these diagrammatic representations can improve syllogistic reasoning.

2.2 Diagrammatic Representations

As emphasized in Stenning and Oberlander (1995), diagrams that are beneficial as a tool in logical reasoning must satisfy two requirements. On one hand, the diagrams used should be simple and concrete enough to express information with their natural and intuitive properties such as geometrical or topological properties. On the other hand, the diagrams used should have some abstraction to deal with the partial information arising in inferential processes. The dilemma here is that in order to manipulate diagrams with abstraction correctly, users need to learn some arbitrary representational conventions governing the abstraction; however the existence of such conventions often clashes with the first requirement, that is, the naturalness of diagrammatic representations. Accordingly, if diagrams are beneficial in syllogistic reasoning, they must have enough abstraction as well as some natural properties exploitable in the reasoning tasks. An Euler diagrammatic representation system introduced in our previous work (Mineshima et al. 2008, 2012a where a formal semantics and a diagrammatic inference system are provided for it), called the EUL system, can serve as such a system (for Stenning and Oberlander's cognitive system of Euler diagrammatic reasoning, see Sects. 2.3.1 and 2.3.3). Hence, of the various representation systems based on Euler diagrams, we focus on the EUL system in our experiment. In this section, we will first show the problem with traditional Euler circles (Gergonne's diagrams) and then explain the Venn diagrams, the modified Euler diagrams (EUL) and the Linear diagrams, which overcome the problems of Gergonne's diagrams.

2.2.1 Problems with Traditional Euler Circles (Gergonne's Diagrams)

Leonhard Euler (1768) introduced circle diagrams to represent syllogistic reasoning with the categorical sentences containing basic quantifiers of *all*, *some*, *some-not*, and *no*.[3] There are several types of diagram that have been studied in Euler's original work (cf. Hammer and Shin 1998) as well as in the subsequent development of diagrammatic logic (for a survey, see Stapleton 2005; Rodgers 2014). Among them, there is a particular version of Euler representation systems, which we call Gergonne's diagrams (Gergonne 1817), a system which is well known in the context of the psychological studies of syllogistic reasoning (e.g., Erickson 1974). This system assumes that diagrams consist only of circles and holds *The Existential Assumption for minimal regions*

[3] Euler invented his diagrams to teach Aristotelian syllogistic logic to a German princess. Indeed, the origin of such diagrams may go back still further. According to Baron's (1969) historical review, we can find the original idea at least in the thirteenth century scholar Ramon Lull (1617). Furthermore, in Leibniz's (1903/1988) work in the seventeenth century, there is the description of the use of diagrams to represent syllogisms although it was only much later that his work was published.

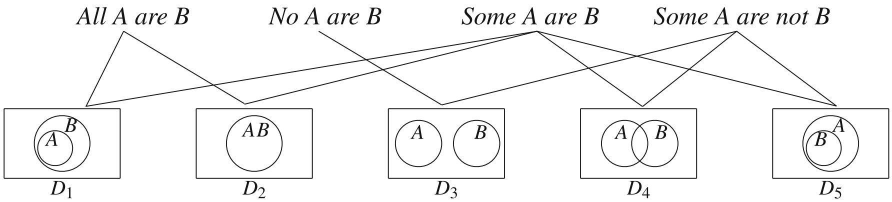


Fig. 2 Correspondence between categorical sentences and Gergonne's diagrams

(EA): every minimal region (namely, a region having no other region contained within it) in a diagram represents a non-empty set. As a consequence, there are five possible combinations for representing the relationship between two sets: see the diagrams in Fig. 2. These are the so-called "Gergonne relations" (cf. Kneale and Kneale 1962, 349–352). It then follows that one categorical sentence may correspond to more than one diagram. Thus, a categorical sentence *All A are B* requires two diagrams, the diagrams D_1 and D_2 as shown in Fig. 2. Similarly, *Some B are C* requires four diagrams of the form, D_1, D_2, D_4, or D_5 in Fig. 2. This means that to check the validity of a syllogism with these two premises, one has to take into account eight ways of combining diagrams, thus giving rise to the problem of "combinatorial explosion" of premise diagrams as emphasized by Johnson-Laird (1983, chap. 4) (see also Roberts et al. 2001). We can see that although Gergonne's diagrams are relatively intuitive in that the semantic convention governing them is solely captured by Existential Assumption (EF), they lack a structured mechanism to deal with abstraction, an essential mechanism to handle logical reasoning. It can be concluded that Gergonne's system has difficulty in handling partial or indeterminate information in a single diagram and hence are difficult to handle in actual logical reasoning. The Euler circles used in the experiments of Rizzo and Palmonari (2005) and Calvillo et al. (2006), as mentioned in Sect. 1, can also be considered to be based on this version of the Euler's representation system. In their experiments, positive results on the efficacy of Gergonne's diagrams were not shown. Given the findings, there is no call for further empirical studies of such diagrams and, after this paragraph, we will focus on the diagrams avoiding this defect.

2.2.2 Venn Diagrams

Venn (1881) and Peirce (1897) attempted to overcome the shortcomings of Gergonne's diagrams by abandonining the Existential Assumption for minimal region (EA) and instead adopting *The Existence-Free Assumption for minimal regions* (EFA): Every minimal region in a diagram is free from the existential assumption: that is, it may denote an empty or non-empty set. Accordingly, the diagram in which circles partially overlap each other as in D_4 of Fig. 2, *viz*, a diagram called "primary diagrams" by Venn (1881), is semantically vacuous, that is, deliver no semantic information about the relation between the sets denoted by the circles. Accordingly, this form of diagram can be used to express that the semantic relationship between the circles is *indeterminate*. Then, in Venn diagrams, meaningful relations among sets are expressed using novel

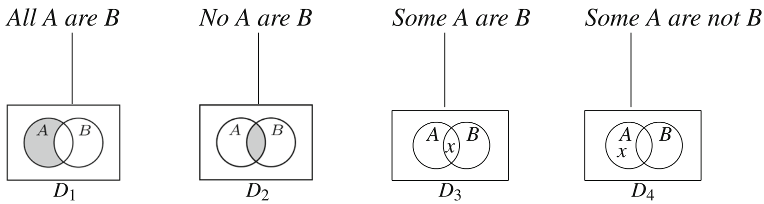


Fig. 3 Correspondence between categorical sentences and Venn diagrams

syntactic devices. To begin with, set inclusion and exclusion relations are expressed using *shading*, by the stipulation that shaded regions denote empty sets. Thus, using negation, the categorical sentences *All A are B* and *No A are B* are first paraphrased as *There is nothing which is A but not B* and *There is nothing which is A and B*, respectively, and then represented as diagrams D_1 and D_2 of Fig. 3.

In this way, logical relations among terms are represented not simply by topological relations between circles, but by the essential use of shading. Then, for graphically representing existential sentences, a *point*, such as x, can be used to indicate the existence of an object in a region. Given this convention, categorical sentences *Some A are B* and *Some A are not B* can be represented using diagrams D_3 and D_4, respectively, in Fig. 3. After all, we can obtain one-to-one correspondence between categorical sentences and Venn diagrams, as shown in Fig. 3.

Venn diagrams are considered to be relatively expressive, and the representation system of Venn diagrams augmented with disjunction is equivalent to monadic first-order logic (Shin 1994). However, compared with Euler diagrams, the way Venn diagrams represent categorical sentences is more involved, in that set inclusion and exclusion are depicted *indirectly* using shaded regions to denote empty sets.

2.2.3 Euler Diagrams EUL

One possible refinement of Venn diagrams can be obtained by using topological relations between circles instead of shaded regions while preserving the Existence-Free Assumption for minimal regions (EFA). This gives us the revised diagrammatic representation system for Euler diagram, a system called the EUL system introduced in Mineshima et al. (2008) and Mineshima et al. (2012a). This system is a representation system combining the features of the original version of Euler circles (Gergonne's diagrams) and Venn diagrams.[4] In the EUL system, universal sentences, *All A are B* and *No A are B*, are represented by D_1 and D_2 in Fig. 4. Thus, following traditional Euler circles (Gergonne's diagrams), the EUL system represents universal sentences in terms of the spatial relations between circles, that is, inclusion and exclusion relations, without using a conventional device such as shading. Existential sentences, *Some A*

[4] Here, "Euler" diagrams refers to diagrams based on topological relations, such as inclusion and exclusion relations, between circles. Thus, both diagrams in Gergonne's system and those in our EUL system are instances of Euler diagrams, whereas Venn diagrams are not. In fact, the Euler diagrams currently studied in diagrammatic logic are typically based on the Existence-Free Assumption for minimal regions (EFA) rather than the type used in Gergonne's system (cf. Stapleton 2005).

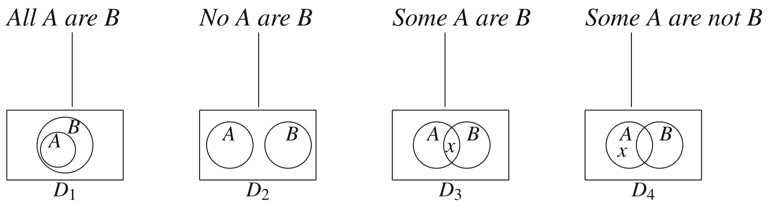


Fig. 4 Correspondence between categorical sentences and Euler (EUL) diagrams

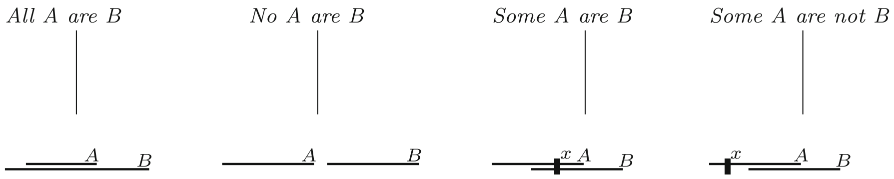


Fig. 5 Correspondence between categorical sentences and Linear diagrams

are B and *Some A are not B*, are represented by D_3 and D_4 in Fig. 4. Thus, we have one-to-one correspondence between categorical sentences and diagrams, a desirable property for an effective use of diagrams in syllogistic reasoning. In summary, the EUL system is distinctive in that it avoids the combinatorial complexities inherent in Gergonne's system. In addition, it dispenses with a new conventional device to express negation, such as shading in Venn diagrams. A common feature of the Venn system and the EUL system is that both rely on the Existential Assumption for minimal region (EA). In what follows, we refer to diagrams in the EUL system simply as Euler diagrams.

2.2.4 Linear Diagrams

A set can be represented not only by circles in a plane but also by a one-dimensional line. Such a linear variant of Euler diagrams is called *linear diagrams*.[5] In linear diagrams, the four kinds of set-relationships expressed by categorical sentences are represented as shown in Fig. 5. Note that in the same way as Euler (circle) diagrams, partially overlapping lines, i.e., those which correspond to sentences *Some A are B* and *Some A are not B* in Fig. 5, indicate that the relationship between the sets represented by these lines is indeterminate, in accordance with EFA. This allows us to obtain one-to-one correspondence between categorical sentences and linear diagrams.

To sum up, we saw that the traditional Euler system ("Euler circles") has fatal problems. Instead, we will use the Euler representation system extended with points (EUL system), a system that gets rid of the problems of the traditional Euler circles.

[5] Linear diagrams for categorical syllogisms were introduced by Leibniz (1903/1988) in the seventeenth century (cf. Politzer et al. 2006). In the later eras, the linear diagrams were developed by the work of Lambert (1764). More recently, Englebretsen (1992) provided a logical system of deductive inference with linear diagrams with two-dimensional spaces. Although it is known that Englebretsen's diagrams are unable to express complex logical inferences and hence have limited expressive power (cf. Lemon and Pratt 1998), linear diagrams are generally expressive enough to represent basic categorical syllogisms.

We will compare this version of Euler diagrams with Venn diagrams. As we will see later, this comparison enables us to distinguish tasks concerning interpretation from those concerning inference so that the existence of syntactic manipulation of diagrams can be tested in an effective way.

2.3 Diagrammatic Inferences

As we saw in Sect. 2.1, there are two ways in which a diagram can contribute to the task of solving deductive reasoning tasks: namely, (a) diagrams can be a clue to supply a correct interpretation to a given sentence and (b) they can be the object of syntactic manipulation to draw a correct conclusion. Section 2.3.1 is devoted to the role of diagrams in (a); we examine what semantic information is associated with categorical sentences when the diagrams we have seen so far are externally given to reasoners (Sect. 2.3.1). Then in Sects. 2.3.2 and 2.3.3, we will see how the manipulation of Euler diagrams help to solve syllogistic reasoning tasks. The case of Venn diagrams is addressed in Sect. 2.3.4.

2.3.1 Semantic Interpretation of Categorical Sentences Imposed by Diagrams

When a categorical sentence is associated with an Euler diagram, a particular meaning is imposed on the sentence in terms of spatial relations hold on the diagram. The relationship between sentence types, diagram types, and their set-theoretical meanings are summarized in Table 1. Here the meaning (truth-condition) of a quantified sentence is specified in term of the relationship between sets, following the standard treatment of *generalized quantifier theory* (Barwise and Cooper 1981) and our previous work (Mineshima et al. 2012b; Mineshima et al. 2014). For example, sentence *All A are B* means that the set denoted by A is a subset of the set denoted by B (symbolically, $\mathbf{A} \subseteq \mathbf{B}$). Here, the quantifier *all* corresponds to the subset relation between sets. Similarly, sentence *No A are B* means that the set denoted by A is disjoint from the set denoted by B (symbolically, $\mathbf{A} \cap \mathbf{B} = \emptyset$), where the quantifier *no* corresponds to the disjointness relation. These set-theoretical meanings can be directly read off from the spatial relationships holding between circles. Thus, the subset relation can be read

Table 1 Semantic interpretations of sentences imposed by Euler diagrams

Sentence	Euler diagram	Spatial relations holding on a diagram	Set-theoretical meaning
All A are B	D_1 in Fig. 4	Circle A is inside circle B	$\mathbf{A} \subseteq \mathbf{B}$
No A are B	D_2 in Fig. 4	Circle A is separated from circle B	$\mathbf{A} \cap \mathbf{B} = \emptyset$
Some A are B	D_3 in Fig. 4	Point x is inside circles A and B	$\mathbf{A} \cap \mathbf{B} \neq \emptyset$
Some A are not B	D_4 in Fig. 4	Point x is inside circle A and separated from B	$\mathbf{A} - \mathbf{B} \neq \emptyset$

Bold type (e.g., **A**) refers to a set denoted by sentence (circle) A

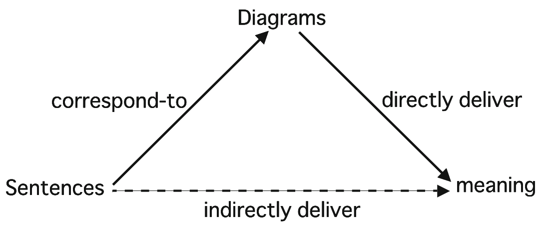


Fig. 6 The relationship between sentences, diagrams, and meanings

off from the fact that circle A is inside circle B as shown in diagram D_1 in Fig. 4; similarly, the disjointness relations can be directly read off from the fact that circle A is separated from circle B as in diagram D_2 in Fig. 4.

What is interesting about our Euler diagrams is that the meaning of existential sentences is also represented in terms of spatial relations, or more specifically, inclusion and exclusion relations, between circles and points. Thus, *Some A are B* means that there is an individual x that is an element of set $\mathbf{A}$ and set $\mathbf{B}$; in other words, the intersection of $\mathbf{A}$ and $\mathbf{B}$ is non-empty (i.e. $\mathbf{A} \cap \mathbf{B} \neq \emptyset$). This semantic information can be read off from diagram D_3 in Fig. 4, where the point x is located inside circle A and circle B. Similarly, *Some A are not B* means that there is an individual x that is an element of set $\mathbf{A}$ but not an element of set $\mathbf{B}$; in other words, the set difference of $\mathbf{A}$ and $\mathbf{B}$, namely, $\mathbf{A} - \mathbf{B}$, is non-empty. This information is exactly what is suggested by the diagram D_4, where the point x is located inside A but outside B.

The role of diagrams in fixing the correct (intended) interpretation of a sentence can be shown as in Fig. 6. One can directly read off the information from a presented diagram, since there is a correspondence between the spatial relations holding on a diagram and the set-theoretical meanings. Categorical sentences associated with Euler diagrams can indirectly communicate the correct information. What one has to do in understanding a given sentence is simply to perceive the associated diagram and to read off the relevant semantic relationship intended by the use of the sentence.

Throughout the paper, we assume that the subject term A of universal sentence *All A are B* does not have *existential import*, that is, it does not imply that there is an individual x in set $\mathbf{A}$. As is well known, this accords with the modern interpretation of quantified sentences but not with the Aristotelian interpretation where *All A are B* implies *Some A are B*. We take it that in ordinary conversation, existential import is derived as a pragmatic interpretation rather than the semantic (truth-conditional) interpretation of a sentence (see Geurts 2003, 2007 for discussion). In our experiments, the participants were instructed to interpret categorical sentences in this way. This choice of the interpretation of universal sentences depends on the choice of diagrams shown in Fig. 4. But note that if we choose to adopt the traditional Aristotelian interpretations, we can instead use Euler diagrams in which some additional points are put in relevant regions. Indeed, Stenning and Oberlander's (1995) cognitive system of Euler diagrams adopts the interpretation that has existential import. In their system, for example, the region inside circle A in diagram D_1 in Fig. 4 has a check mark; the regions inside circle A and circle B in diagram D_2 in Fig. 4 have check marks respectively (the subject-predicate order of each conclusion is both CA and AC). Given these diagrams, categorical sentences should be interpreted as having existential import. In

Stenning and Oberlander's account, however, check mark play an essential role in the process of solving syllogisms (for more details, see Sect. 2.3.3). Thus, their account is inevitably limited to the case of syllogisms having a categorical interpretation with existential import (for a similar discussion, see Boolos 1984). By contrast, our system has a broader scope, because it separates the issue of existential import from the core mechanism of (in)validity checking. Our diagrams can be applied to interpretations with and without existential import and thus can be used for interpretation control in logical reasoning tasks.

If the semantic information encoded by categorical sentences was directly accessible to untrained users even when diagrams were not externally given, it would be some easier for them to solve categorical syllogisms. However, cognitive psychological studies of logical reasoning accumulated so far suggest that this is not the case. The following two classes of "interpretation errors" are particularly worth focusing on. Firstly, untrained people tend to take inconvertible pairs of terms as convertible. For example, *All A are B* is often interpreted as equivalent to sentence *All B are A*. This is what is traditionally called *conversion errors* in the literature (Chapman and Chapman 1959). Conversely, untrained people can interpret convertible pairs of terms as inconvertible; thus, one may fail to notice that sentence *Some A are B* is semantically equivalent to sentence *Some B are A*. This suggests that the subject-predicate structure of a sentence is responsible for difficulties in obtaining a correct interpretation of a categorical sentence. Such factors concerned with word order have been justly regarded as ones causing *figural effects* (e.g., Dickstein 1978). One can avoid these two types of "interpretation errors" by perceiving the information provided by Euler diagrams. A more detailed analysis and a particular prediction provided by our account will be presented in Sect. 2.4.2.

Before moving on to discussion on inferential efficacy of Euler diagrams, we emphasize that such interpretations as indicated above are not genuine errors but rather reasonable ones that can be derived from pragmatic considerations. For instance, the inference from *All A are B* to *All B are A* is an instance of *invited inference* or *pragmatic perfection* (Geis and Zwicky 1971). Such an inference is standardly treated as *scalar implicature* in pragmatic theory (Horn 2000), and hence, not due to the semantics of universal quantifiers. Also, the subject-predicate structure of sentences like *Some A are B* can be regarded as an instance of more general phenomena concerned with topic-focus structure or information structure of sentences (cf. Stenning 2002). Generally speaking, diagrams are effective in fixing the intended interpretation of sentences used in reasoning tasks, not because they deliver the "correct" interpretations of sentences, but because they succeed in abstracting away from pragmatic aspects of meaning and hence in extracting the purely semantic information from a given sentence. This means that external diagrams can *control* experiments on deductive reasoning tasks in that we can fix the intended interpretation of a sentence and thereby separate between interpretational and inferential aspects of the process of solving reasoning tasks. In this sense, the use of diagrams may shed light on deductive reasoning in general, contributing to a more fine-grained design of experiments concerning deductive reasoning.

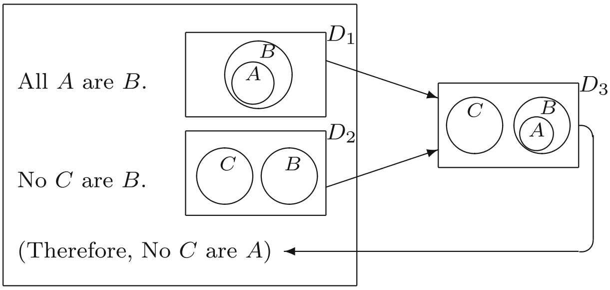


Fig. 7 A solving process of a syllogism (AE2E type) using Euler diagrams

2.3.2 Reasoning with Euler Diagrams: Single Type

In the previous section, we saw how Euler diagrams function as a visual representation of the intended set-theoretical meaning of premise sentence. Such a representation can also be an effective vehicle of inference in solving deductive reasoning tasks. In this section, we will see how Euler diagrams can be manipulated to solve syllogistic reasoning tasks. This is based on a formal system of Euler diagrammatic inference in Mineshima et al. (2008) and Mineshima et al. (2012a) where soundness and completeness are provided for it.[6]

As an illustration, consider the example shown in Fig. 7, which shows the process of solving the syllogism given in Fig. 1 at Sect. 1. Suppose that a reasoner is presented with two categorical sentences and corresponding Euler diagrams, D_1 and D_2, as indicated in the box. The overall process of solving a syllogism with Euler diagram can be divided into three phases, which we call *interpretation*, *manipulation*, and *recognition*. First, in the interpretation phase, a reasoner extracts the intended meaning of premise sentences from the spatial relationships holding on diagrams. Thus, in this example, the reasoner observes the *inclusion* relation holding between circles A and B in diagram D_1 and interprets the sentence *All A are B* as meaning that the set denoted by A is a subset of the set denoted by B; similarly, the reasoner can interpret the sentence *No C are B* as meaning that the set denoted by C is disjoint from the set denoted by B by observing the *exclusion* relation holding between circles C and B in the diagram D_2. Second, in the manipulation phase, the reasoner combines the information contained in given premises. Given an association of premise sentences with Euler diagrams, such a process of combining information can naturally be replaced with a process of syntactic manipulation diagrams. More specifically, the manipulation phase consists of processes of *unifying* premise diagrams. In this case, the reasoner can unify the diagrams D_1 and D_2 to obtain the diagram D_3 in Fig. 7. The strategy of manipulating diagrams is simple enough: one identifies the circle B in the premise diagrams, while preserving the other relations holding with respect to circles A and B. To be more precise, there are at least two ways of unifying diagrams: incorporating the circles of D_1 into D_2 or vice versa—both ways yield the same result. This unification process

[6] The EUL system can derive not only simple syllogisms with two premises but also a chain of syllogisms with more than two premises (the so-called *sorites*).

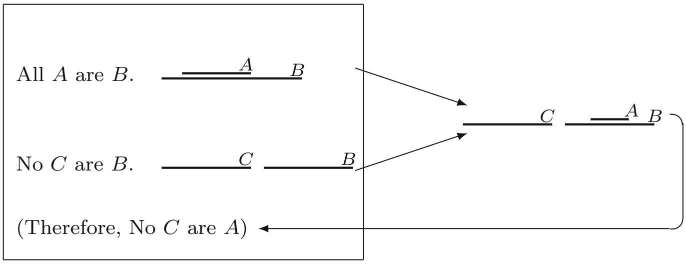


Fig. 8 A solving process of a syllogism (AE2E type) using Linear diagrams

almost automatically determines the relation between circles A and C: without efforts, circle A is located outside circle C. Finally, in the recognition phase, observing that the exclusion relation holds between circles A and C, one can infer that the set denoted by A is disjoint from the set denoted by C and thus successfully draw the conclusion that *No C are A*.

A syllogistic reasoning task can be solved by unifying linear diagrams associated with premise sentences, as in the case of Euler diagrams. As an example, consider the case of a syllogism *All A are B*, *No C are B*; *therefore, No C are A*. The unification process is shown in Fig. 8, which corresponds to the case with Euler diagrams shown in Fig. 7. By identifying the line B in the premise diagrams, one can automatically obtain the linear diagram in which the line C is separated from the line A, from which one can draw the valid conclusion that *No C are A*. The solving processes of the other cases are essentially the same as the ones in Euler diagrams.

It is important to see that the way in which two diagrams are unified is *spatially* constrained; that is, the spatial relations holding between circles in premise diagrams determines possible ways of unification. More specifically, there are two constraints at work here: for any circle or point X and for any circle Y and Z, (1) if X is inside Y in one diagram D_1 and Y is inside Z in another diagram D_2, then X is inside Z in the combined diagram $D_1 + D_2$; (2) if X is inside Y in one diagram D_1 and Y is separated from Z in another diagram D_2, X is separated from Z in the combined diagram $D_1 + D_2$. It can be easily seen that the constraint (2) is responsible for the correct manipulation in the case of Fig. 7. Note that as summarized in Table 1, spatial relations (inclusion and exclusion relation) holding on diagrams correspond to set-theoretical relations (subset and disjointness relations, respectively). Based on this correspondence, one can draw a semantically correct information in terms of the process of unification. Specifically, the constraint (1) is amount to saying that if **X** is a subset of **Y** and **Y** is a subset of **Z**, then **X** is a subset of **Z**, and the constraint (2) to saying that if **X** is a subset of **Y** and **Y** is disjoint from **Z**, then **X** is disjoint from **Z**. We can see that in this sense, reasoning about semantic information (set-theoretical relations) can be replaced by reasoning about spatial relations holding on diagrams.

An important characteristic of the unification process in diagrammatic reasoning is that by combining two premise diagrams, one can almost automatically determine the semantic relation between the objects in question, without any additional operation. Shimojima (1996) refers to such information that can be automatically inferred from the result of a diagrammatic operation as "free ride" (cf. for a partial formal specifica-

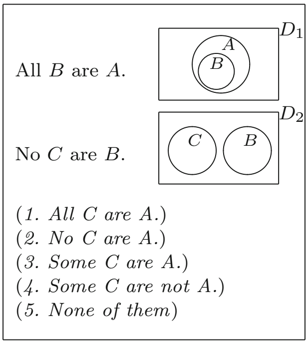


Fig. 9 An example of a syllogism with no valid conclusion: the correct answer is 5

tion of free ride property in logical reasoning with the Euler diagrams, see Takemura 2013). The constraints (1) and (2) are purely spatial ones, so that other possible ways of unification given the same premise diagrams would not be available to users. Moreover, the process of unification is so simple/natural and effortless that it is available to even people without training of manipulating diagrams (for empirical supports from experiments using nesting cup tasks, see Greenfield et al. 1972; Deloache et al. 1985; Johnson-Pynn et al. 1999 and for more general perspective of embodiment cognition, see Johnson 1987, chap. 2, and Lakoff and Nunez 2000, chap. 6). We would expect that even untrained reasoners could manipulate Euler diagrams in a suitable way without much efforts and succeed in drawing a valid conclusion.

A characteristic of cases like Fig. 7 is that the spatial relation between objects (circles and points) is uniquely determined via unification process. Such cases will be called *single* type. A complication arises when one consider a case in which unification process does not uniquely determine the spatial relation between objects—a case that will be called a *multiple* type. Now we will turn to examining how Euler diagrams can be effective in such multiple unification type.

2.3.3 Reasoning with Euler Diagrams: Multiple Type

A typical example of syllogisms involving multiple type is one in which no (non-trivial) valid conclusion can be deduced from premises. A task to judge that a given pair of syllogistic sentences has no valid conclusions will be called an *NVC task*. An NVC task is known to be particularly difficult to solve for untrained people (cf. Revlis 1975; Khemlani and Johnson-Laird 2012).[7] Hence it is interesting to see whether Euler diagrams can be effective to solve NVC tasks.

As an example, consider a syllogism with two premises, *All B are A* and *No C are B*, as shown in Fig. 9. A solving process runs as follows. First, in the interpretation phase, the diagram D_1 where circle B is included in circle A triggers the interpretation that the set denoted by B is a subset of the set denoted by A, and the diagram D_2 where

[7] Furthermore, Stenning et al. (1995) reported that diagrams are less effective in the NVC tasks when leaning logic with a computer-assisted system introduced by Barwise and Etchemendy (1994), *Hyperproof*, using a hybrid interface of logical formulas and diagrams.

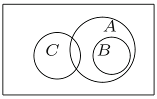


Fig. 10 The unified diagram for the syllogism in Fig. 9 in accord with the semantics of Euler diagrams

circle C is separated from circle B triggers the interpretation that the sets denoted by B and C are disjoint. In the manipulation phase, then, the reasoner tries to unify the diagrams D_1 and D_2. Here, one can identify circle B in two diagrams but fails to determine the unique spatial relations with respect to circles A and C; that is, neither inclusion nor exclusion relations can be inferable from the information contained in the two premises. From this fact, one can correctly draw the inference that there is no valid conclusion in this case.[8] What plays a crucial role here is the *specificity* of diagrammatic representation, which forces the reasoner to put diagrammatic objects in a specific way when combining two diagrams (cf. Stenning and Oberlander 1995). The specificity of diagrams often impedes reasoning, resulting in what Shimojima (1996) calls the *over-specificity* of diagrammatic representations. In the present case, however, the specificity of diagrams has a positive effect in that it supports the process of checking the invalidity of a syllogistic inference. That is, a failure to determine the unique position of diagrammatic objects, such as circles A and C in the above case, could trigger the recognition that a given set of premises does not have a (non-trivial) valid conclusion.

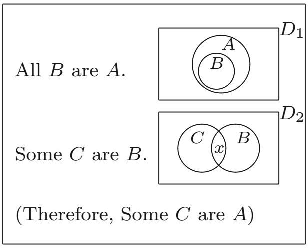
It is noted here that one can unify the two diagrams in Fig. 9 in accord with EFA discussed in Sect. 2.2.2. It gives rise to the diagram shown in Fig. 10. Here the fact that circle A partially overlaps with circle C indicates that no specific semantic information about terms A and C can be drawn from the premises.[9] It should be noted that to put the final unified diagram in this way is not a necessary condition to select the correct answer (that is, to judge that there is no valid conclusion); to solve an NVC task with Euler diagrams, it is *sufficient* to observe that unification does not lead to inclusion nor exclusion relation with respect to the circles (or points) in question. A sophisticated reasoner with full understanding of the semantics of Euler diagrams can construct the diagrams in Fig. 10, but even those who lack such full understanding could solve NVC tasks successfully.

Multiple type can also arise in a syllogism with a valid conclusion. More specifically, a syllogism having an existential premise involves multiple unification. An example is given in Fig. 11. In this case, by identifying circle B one can automatically infer, under the constraint (1), that point x is inside circle A, in a similar way to the case shown in Fig. 7 above. With respect to circles A and C, neither inclusion nor exclusion relations is inferable from the information contained in the two premises. However, in

[8] It is noted that this process of checking the invalidity of inferences is similar to the one known as "negation as failure" in the AI literature. One can see that the role of diagrams is to limit the possible search space of drawable conclusions.

[9] The relevant constraint on unification, called the constraint (3), is the following: for any circles X, Y and Z, if neither inclusion or exclusion relation holds between X and Y in the combined diagram, put X and Y in such a way that X and Y are partially overlapped each other.

Fig. 11 A syllogism (AI1I type) with an existential premise that involves multiple unification



either case, one can observe that there is a point x inside circles A and C, and hence can correctly draw the conclusion that some C are A.

Cases in which multiple type is involved are relatively complex in that the reasoner is forced to enumerate more than one possibilities. Accordingly, it is expected that solving multiple unification cases like Figs. 10 and 11 requires more efforts than single unification cases like Fig. 7.

The multiple type makes explicit the difference between our system and Stenning and Oberlander's (1995) cognitive system of Euler diagrammatic inference (for their representation system of Euler diagrams, see Sect. 2.3.1). In Stenning and Oberlander's system, the premise diagrams are unified in such a way that the conclusion diagrams represent maximal models, a convention called *crossing circles* in our study, and users who do not precisely follow this convention inevitably fail to solve the reasoning tasks with their Euler diagrams. As stated the above, by contrast, our system allows users to choose the strategy of crossing circles as well as the strategy that does not depend on the full understanding of crossing circles. This flexibility can make it easier for users to unify the premise diagrams in the multiple type. In Stenning and Oberlander's system, furthermore, there is a conventional rule about check marks during the unification of premise diagrams. If a region with a check mark is divided into smaller regions, then the check mark is removed. This plays an essential role in the process of solving syllogisms in their system. Thus users who do not precisely follow this convention inevitably fail to reason with the Euler diagrams.

These conventional rules seem to be artificial in that they are not available to users without training in diagram manipulation. Thus, the possible processes of diagrammatic reasoning which were proposed in Stenning and Oberlander (1995) are not expected to have a significant positive effect, compared to reasoning with other forms of representations. Indeed, several studies of logic teaching with Stenning and Oberlander's Euler diagrams seem to support this view. Monaghan and Stenning (1998) taught students how to manipulate Stenning and Oberlander's Euler diagrams to solve syllogisms, but they found no significant difference between results of teaching with the Euler diagrams and teaching with symbolic formulas (as in a natural deduction system). Dobson (1999) examined the effects of learning software with a graphical interface of Stenning and Oberlander's Euler diagrams in solving syllogistic tasks. In contrast to the predictions arising from his theoretical analysis, the Euler software was less effective than other graphical software using Venn diagrams.

It should be noted that Monaghan and Stenning (1998) emphasized participants' "individual differences" in diagrammatic reasoning and learning and found advantages for reasoning using Stenning and Oberlander's (1995) Euler diagrams. Their study revealed that effective users of the Euler system had high ability for spatial manipulation and used a holistic learning strategy. Some conventional devices in Stenning and Oberlander's system may be difficult for most users to understand but may not trouble a small segment of users who may view the Euler method as a natural system of mechanical and spatial manipulation (for a discussion of the "mechanical" characteristics of Euler diagrammatic reasoning, see Stenning 2002, chap. 4). Although Stenning and Oberlander's system has important implications for actual (educational) applications of diagrammatic reasoning, we focus on fundamental aspects of diagrammatic reasoning efficacy, based on an analysis of Sect. 2.1.

2.3.4 Reasoning with Venn Diagrams

One possible way to solve the syllogism using Venn diagrams is illustrated in Fig. 12, where the premise *All B are A* is represented by D_1, and the premise *No C are B* by D_2 is externally presented. Here two premise diagrams have different sets of circles. In such a case, in accord with the syntax of Venn diagrams, the configurations of circles must be accommodated by adding circle C to D_1, and circle A to D_2. Then, by superposing the shaded region of D_3 on D_4, one can obtain diagram D_5, from which the correct conclusion *Some C are A* can be reached. In general, the solving process using Venn diagrams consists of two steps, referred to as addition and superposition. Note that a process of combining Euler diagrams, namely *unification*, can exploit the movements of circles as in Fig. 7, whereas a process of combining Venn diagrams, namely *superposition*, operates on premise diagrams with the same number of circles, and hence does not involve any movement of circles. In order to combining Venn diagrams given as premises, one has to know the relevant inference rules and strategies in advance. More specifically one has to know the successive processes of adding a circle and superposing two diagrams (i.e., the generation of D_3 and D_4). Note that although the operation of superposition would be triggered by the goal of the deduction task itself, the operation of addition is a *prerequisite* for superposition and thus seems to be not triggered directly. We expect that those who are ignorant of such a solving strategy could not appeal to concrete manipulations of the diagrams. They seem to have to draw a conclusion solely based on usual linguistic inference, with the help of semantic information obtained from Venn diagrams.

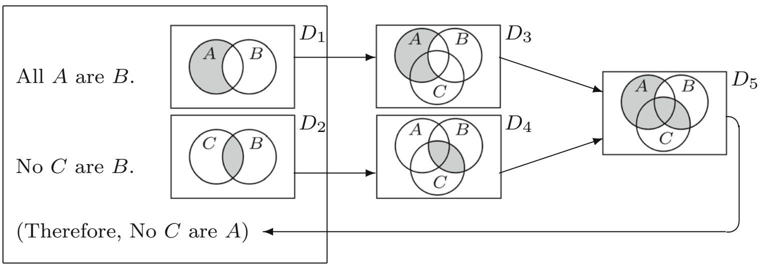


Fig. 12 A solving process of a syllogisms (AE2E type) using Venn diagrams

2.4 Predictions

2.4.1 Predictions of Inferential Efficacy

We claim that Euler diagrams have inferential efficacy, that is, syntactic manipulation of diagrams can occur in logical reasoning with Euler diagrams. One way to test the existence of inferential efficacy would be to compare performance in logical problem solving with distinct diagrams that are equivalent in semantic information but differ in form, specifically, to compare diagrams with forms that are suitable for syntactic manipulation and diagrams with forms that are not. Comparing the effectiveness of Euler diagrams with that of Venn diagrams would be suitable for this purpose. We hypothesized that Euler diagrams provide concrete form to the abstract relational structures of categorical sentences. This enables a reasoner to avoid interpretational errors in understanding categorical sentences, and, more importantly, to replace the task of drawing a conclusion with a more concrete process of manipulating diagrams. Thus we claim that Euler diagrams have not only interpretational efficacy but also inferential efficacy in the sense specified in Sect. 2.1. By contrast, we expect that Venn diagrams convey the relational information contained in categorical sentences indirectly, using the device of shading; thus, Venn diagrams can be effective in avoiding interpretational errors, but the form of Venn diagrams is unsuitable for the process of unification that is essential for syllogistic reasoning tasks, hence they lack inferential efficacy. Based on these considerations, we predict that the performance of solving syllogisms would be better when participants use Euler diagrams than when they use Venn diagrams or when they do not use any diagrams.

When participants use Euler diagrams, there may be difference in ease of unification between syllogisms of single type and those of multiple type; as suggested in Sects. 2.3.2 and 2.3.3, we expect that syllogisms of single type, namely, those whose unification results in the unique configuration of circles or points (cf. Fig. 7) would be easier to perform than syllogisms of multiple type, namely, those which requires the reasoner to consider multiple possibilities (cf. Fig. 11). The syllogisms of single type include AA1, AE2, AE4, EA1, EA2 type syllogisms under the labeling summarized in "Appendix 2". The syllogisms of multiple type are divided into two groups: those with valid conclusion, including AI1, AI3, IA3, IA4, AO2, OA3, EI1, EI2, EI3, EI4 types, and those with no valid conclusion, including AA2, AA3, AI2, AI4, IA1, IA2, AE1, AE3, EA3, EA4, AO1, AO4, OA1, OA4 types. See "Appendix 2" for a list of these syllogisms associated diagrams.[10]

2.4.2 Predictions of Interpretational Efficacy

As noted in Sect. 2.1, in the case of linguistic syllogistic reasoning, where participants are not allowed to use any diagram, it is known that participants often make some interpretational errors due to the word order, such as the subject-predicate distinction,

[10] This classification of one/multiple model syllogisms is different from one in the mental model theory. In contrast to this study, for example, Bucciarelli and Johnson-Laird (1999) classified the AI1 syllogism having the premises *All B are A* and *Some C are B* to a single model syllogism.

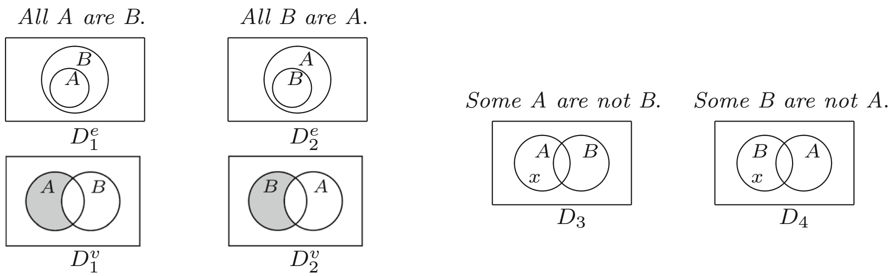


Fig. 13 Sentences that tend to cause conversion errors and corresponding topologically non-identical diagrams that clarify their intended meanings

of a sentential material (cf. Newstead and Griggs 1983). It is expected that both Euler diagrams and Venn diagrams may help participants avoid such interpretational errors in linguistic syllogistic reasoning. In Sects. 2.3.2 and 2.3.3, we divided these interpretation errors into two types, namely, conversion errors and figural effects. For each type, we will explain what class of syllogisms would be improved with the use of diagrams.

Conversion errors As shown by the experiments in Chapman and Chapman (1959) and Dickstein (1981), among others, an illicit conversion of terms appearing in premises may cause errors in syllogisms. For example, in the case of AE1 syllogism, the premise *All B are A* tends to be interpreted as equivalent to *All A are B*, leading the reasoner to select the incorrect answer *No C are A*. When the reasoner is presented with Euler diagrams, the two sentences *All A are B* and *All B are A* correspond to D_1^e and D_2^e of Fig. 13, respectively. Here, one could immediately see that these two diagrams are topologically different, and hence, deliver different information. Similarly, in the case Venn diagrams, these two sentences correspond to D_1^v and D_2^v, respectively, and the diagrams could make it clear that the terms in question are not convertible. A similar explanation applies to the pair of *Some A are not B* and *Some B are not A*, which is represented in Euler and Venn diagrams as D_3 and D_4 in Fig. 13. The class of syllogisms that may cause conversion errors consists of AA2, AA3, AA4, AI2, AI4, IA1, IA2, AE1, AE3, EA3, EA4, AO1, AO3, AO4, OA1, OA2, and OA4 types, all being syllogisms with no valid conclusion. Thus, the use of diagrams seems to block the errors caused by the misinterpretation of categorical sentences. We predict that errors causes by illicit conversion are reduced when Euler and Venn diagrams are presented to reasoners.

Figural effects Because of the strict distinction between subject and predicate in categorical sentences, it is sometimes difficult to understand the logical equivalence between the E-type sentences *No A are B* and *No B are A* and also between the I-type sentences *Some A are B* and *Some B are A*. Dickstein (1978) reported that such a difficulty appeared most prominently as a difference in the performances between EI1O and EI4O syllogisms, which have the above sentences as premises (EI1O refers to *No B are A, Some C are B*; therefore *Some C are not A*. EI4O refers to *No A are B, Some B are C*; therefore *Some C are not A*.). He also pointed out that the difference was a notable example of the *figural effect*. In Euler and Venn diagrams, E-type and I-type

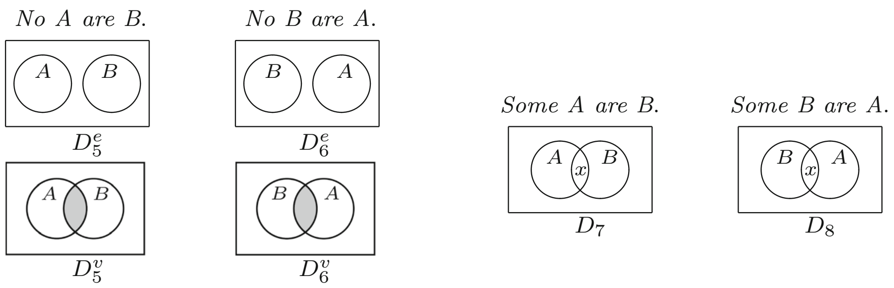


Fig. 14 Sentences that tend to suffer from figural effects and corresponding topologically identical diagrams that clarify their meanings

sentences are represented as shown in Fig. 14. Here, it seems to be easy to understand the equivalence of D_5^e and D_6^e (also of D_5^v and D_6^v, and of D_7 and D_8) since they are topologically identical.

3 Experiment

3.1 Method

We compared participants' performances in syllogism solving in cases where diagrammatic representations (i.e. Euler diagrams and Venn diagrams) are used with the cases where they are not used. Before the reasoning tasks, we provided the participants with instructions on the meanings of diagrams and sentences, but not on particular rules or strategies of how to use diagrams to solve syllogisms. More specifically, as described in Sect. 2.2, the point of the instruction is that a circle is used to denote a set of objects and point x is used to indicate the existence of an object for the Existence-Free Assumption for minimal regions (EFA). The convention of EFA seems technical so that some instructions are needed to understand it correctly. Accordingly, we provided participants with instructions on the meanings of diagrams to fix the intended interpretation of the diagrams used in the experiments.

3.1.1 Participants

Two hundred and thirty-six undergraduates (mean age 20.13 ± 2.99 SD) in five elementary philosophy classes participated in the experiment. They gave their consent to their cooperate in the experiment, and after the experiment, they were given a small non-monetary reward. None had any prior training in syllogistic logic. The participants were native speakers of Japanese, and the sentences and instructions were given in Japanese.[11] The participants were randomly divided into three groups: the

[11] We used the following translation: "Subete no A wa B de aru" for *All A are B*, "Dono A mo B de nai" for *No A are B*, "Aru A wa B de aru" for *Some A are B*, "Aru A wa B de nai" for *Some A are not B*. Here we use the quantifiers "subete" and "dono" for *all*, and "aru" for *some*. One remarkable difference between English and Japanese is in the translation of *No A are B*. Since in Japanese there is no negative

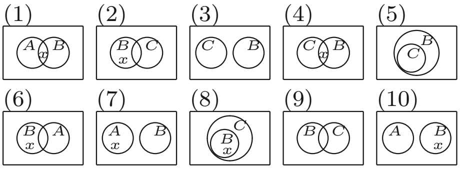


Fig. 15 Euler diagrams used in the pretest

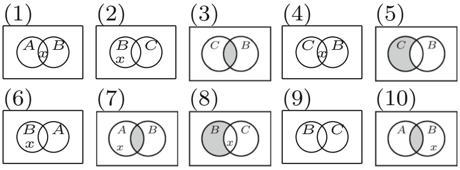


Fig. 16 Venn diagrams used in the pretest

Linguistic group, the Euler group, and the Venn group. The Linguistic group consisted of 66 participants. Of them, we excluded 21 participants: those who left the last more than three questions unanswered (19 participants)[12] and those who had participated in our pilot experiments conducted before (2 participants). The Euler group consists of 68 participants. Of them, we excluded 5 participants: those who left the last more than three questions unanswered (3 participants) and those who had participated in our pilot experiments conducted before (2 participants). The Venn group consists of 102 participants. Of them, we excluded 34 participants: those who left the last more than three questions unanswered (27 participants) and those who had participated in our pilot experiments conducted before (7 participants). It was not examined to what extent participants have prior knowledge of each diagram.

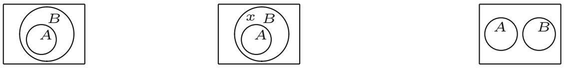
3.1.2 Materials

The experiment was conducted in the booklet form.

Pretest The participants of the Euler group and Venn group were presented with 10 diagrams listed in Figs. 15 and 16, respectively. The participants were asked to choose, from a list of five possibilities, the sentences corresponding to a given diagram. Examples are given in Figs. 17 and 18. The answer possibilities were *All*-, *No*-, *Some*-, *Some-not*, and *None of them*. The subject-predicate order of an answer sentence was AB or BC. The total time given was five minutes. The correct answer to each diagram was: (1) *Some A are B*, (2) *Some B are not C*, (3) *No B are C*, (4) *Some B are C*, (5) *None of them*, (6) *None of them*, (7) *No A are B* and *Some A are not B*, (8) *All B are C* and *Some B are C*, (9) *None of them*, (10) *No A are B*, respectively. The highest possible score on the pretests of the Euler, and Venn groups was twelve, because there were two correct answers in two of the ten problems. The total time in Euler and

Footnote 11 Continued
quantifier corresponding to *No*, we use the translation "Dono A mo B de nai", which literally means *All A is not B*. Except this point, we see no essential differences between English and Japanese. So we will refer to English translation in this paper.

[12] They were regarded as the participants who gave up halfway and dropped out.



1. All A are B.
2. No A are B.
3. Some A are B.
4. Some A are not B.
5. None of them.

Correct answer: 1

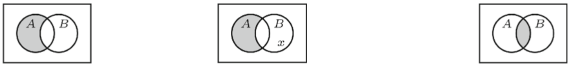
1. All A are B.
2. No A are B.
3. Some A are B.
4. Some B are not A.
5. None of them.

Correct answer: 1 & 4

1. All A are B.
2. No A are B.
3. Some A are B.
4. Some A are not B.
5. None of them.

Correct answer: 2

Fig. 17 The examples in the pretest of the Euler diagrams



1. All A are B.
2. No A are B.
3. Some A are B.
4. Some A are not B.
5. None of them.

Correct answer: 1

1. All A are B.
2. No A are B.
3. Some A are B.
4. Some B are not A.
5. None of them.

Correct answer: 1 & 4

1. All A are B.
2. No A are B.
3. Some A are B.
4. Some A are not B.
5. None of them.

Correct answer: 2

Fig. 18 The examples in the pretest of the Venn diagrams

Venn groups was 5 min. Before the pretest, the participants were presented with the examples in Figs. 17 and 18.

Syllogistic reasoning tasks The participants in the Euler group were given syllogisms with Euler diagrams (such as the one in Fig. 19). The participants in the Venn group were given syllogisms with Venn diagrams (such as the one in Fig. 20), and participants in Linguistic group were given syllogisms without diagrams. The participants were presented with two premises and were asked to choose, from a list of five possibilities, a sentence corresponding to the correct conclusion. The list consists of *All*-, *No*-, *Some*-, *Some-not*, and *NoValid*. The subject-predicate order of each conclusion was CA for the limitation of the number of choices. We gave 31 syllogisms in total, out of which 14 syllogisms had a valid conclusion and 17 syllogisms had no valid conclusion (14 syllogisms had no valid conclusion in both CA and AC orders; 3 syllogisms had no valid conclusion only in CA order). The test was a 20-min power test, and each task was presented in random order (10 patterns were prepared). Before the test, the example in Fig. 19 was presented to participants in the Euler group, and the one in Fig. 20 to participants in the Venn group.

3.1.3 Procedure

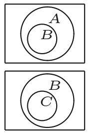
All three groups were first given 1 min 30 s to read one page of instructions on the meaning of categorical statements.[13] In addition, the Euler group was given 2 min to read two pages of instructions on the meaning of Euler diagrams, and the Venn group

[13] Here our instruction emphasized that the meaning of categorical sentences used in our experiment does not contain the existential import. Concretely, the following is given: *All A are B* does not imply that there are some objects which are A; thus, *All A are B* does not imply *Some A are B*. Similarly, *No A are B* does not imply that there are some objects which are A. Thus, *No A are B* does not imply *Some A are not B*.

Fig. 19 An example of syllogistic reasoning task of the Euler group

All B are A.

All C are B.



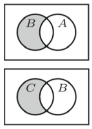
1. All C are A.
2. No C are A.
3. Some C are A.
4. Some C are not A.
5. None of them.

Correct answer: 1

Fig. 20 An example of syllogistic reasoning task of the Venn group

All B are A.

All C are B.



1. All C are A.
2. No C are A.
3. Some C are A.
4. Some C are not A.
5. None of them.

Correct answer: 1

was given 2 min to read two pages of instructions on the meaning of Venn diagrams. Before the pretest, the Euler and Venn groups were given 1 min 30 s to read two pages of instructions on the pretest. Finally, before the syllogistic reasoning test, all three groups were given 1 min 30 s to read two pages of instructions, in which the participants were warned to choose only one sentence as answer and not to take a note. These time limits were set based upon the results of our pilot experiments.

3.2 Results

3.2.1 Pretest

The accuracy rate of each item in the pretest of Euler diagrams, listed in Fig. 15, was (1) 77.8 %, (2) 77.8 %, (3) 90.5 %, (4) 81.0 %, (5) 69.8 %, (6) 84.1 %, (7) 88.9/57.1 % (49.2 %), (8) 95.2/65.1 % (58.7 %), (9) 79.4 %, and (10) 82.5 %, respectively.[14] The accuracy rate of each item in the pretest of Venn diagrams, listed in Fig. 16, was (1) 85.3 %, (2) 75.0 %, (3) 73.5 %, (4) 82.4 %, (5) 39.7 %, (6) 64.7 %, (7) 76.5/50 % (35.3 %), (8) 63.2/55.9 % (35.3 %), (9) 77.9 %, and (10) 66.2 %, respectively. In our experiment, the average accuracy rate of the pretest for Euler group was significantly higher than the rate for the Venn group (79 % for the Euler group and 68 % for the Venn group) [Mann–Whitney $U = 1502.5$, $P = 0.0029$]. This result suggests that the Euler diagrams are easier to learn than the Venn diagrams having the convention of shading. This is in contrast to Dobson's (1999) report in which the learning software

[14] The accuracy rate for those tasks which have more than one correct answer is described as "X/Y % (Z %)", where X % represents the accuracy rate for the first correct answer, Y % the second, and Z % the rate for those who selected both correct answers.

using Venn diagrams was more effective than that using Euler diagrams.[15] Typical pattern of error in the pretest of Euler and Venn diagrams is that both *Some*- and *Some-not* are incorrectly selected in (1), (2), (4), (5), (6), and (9). This suggests that the Existential Assumption for minimal regions (EA) can be robust in novice learners' interpretation. Thus this means that the understanding the Existence-Free Assumption for minimal regions (EFA) for Venn and Euler diagrams requests learners' effort. In the Euler group, this error was observed in 18 participants (out of 63 participants), who scored less than 8 on the pretest (out of 12). In the following analysis, we exclude these 18 participants and refer to the other 45 participants as the restricted Euler group. In the Venn group, this error was observed in 38 participants (out of 68 participants) who scored less than 8 on the pretest (out of 12). In the following analysis, we exclude these 38 participants and refer to the other 30 participants as the restricted Venn group.

3.2.2 Syllogistic Reasoning Tasks

Figure 21 shows the average accuracy rates of the total 31 syllogisms in the three groups. The rate of the Linguistic group was 46.7 %, the rate for the Venn group was 47.8 %, and the rate for the Euler group was 77.8 %. The results of each syllogistic type are shown in "Appendix 1". These data were subjected to a one-way Analysis of Variance (ANOVA). There was a significant main effect, [$F(2, 173) = 39.076$, $p < 0.001$]. Multiple comparison tests by Ryan's procedure yield the following results: (i) The accuracy rate of reasoning tasks in the Euler group was higher than that in the Linguistic group: 46.7 % for the Linguistic group and 77.8 % for the Euler group [$F(1, 106) = 7.280$, $p < 0.001$]. (ii) The accuracy rate of reasoning tasks in the Euler group was higher than that in the Venn group: 47.8 % for the Venn group and 77.8 % for the Euler group [$F(1, 129) = 7.825$, $p < 0.001$]. (iii) There was no significant difference between the Venn group and the Linguistic group in the accuracy rate of reasoning tasks: 47.8 % for the Venn group and 46.7 % for the Linguistic group [$F(1, 111) = 0.274$, $p > 0.1$]. The rate for the restricted Euler group (excluding those who failed the pretest) was 85.2 % and that for the total Venn group (excluding those who failed the pretest) was 66.8 %. The results of each syllogistic type in the restricted Euler and restricted Venn groups are shown in "Appendix 1".

Detailed performance data of the restricted Euler group are provided in Table 2. In this table, "single" refers to single type syllogisms in unifying premise diagrams and "multiple" refers to multiple type syllogisms in unifying premise diagrams. Furthermore, "VC" refers to syllogisms having valid conclusion, and "NVC" refers to syllogisms having no valid conclusion. It should be noted that, in the following analysis of NVC tasks, three syllogisms (AO3, OA2, and AA4) were excluded because they were special cases that have NVC only in the experimental setting (that is, the

[15] Our result is consistent with those presented in Chapman et al. (2014), where the ease of comprehension of Linear, Euler and Venn diagrams was examined. Sato et al. (2011) also reported a similar result on the information extraction from Euler and Venn diagrams. Their empirical results suggested that the interpretation of Venn diagrams requires more substantial efforts than that of Euler diagrams.

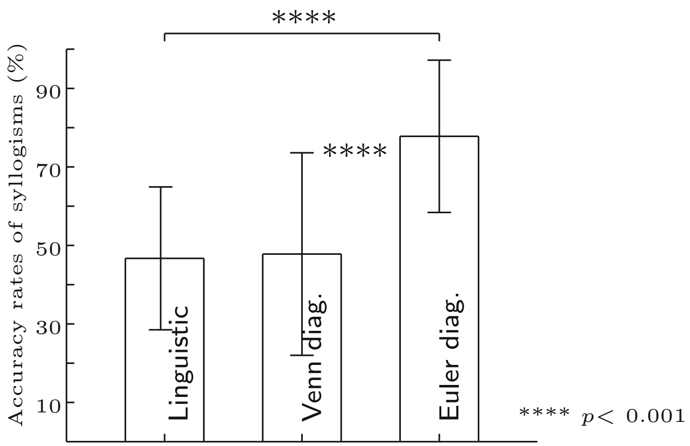


Fig. 21 The average accuracy rates of 31 total syllogisms in the Linguistic group, the Venn group, and the Euler group (*error-bar* refers to SD)

Table 2 The average accuracy rates of the three types in the restricted Euler group (compared to the restricted Venn group)

	Single	Multiple	
	VC (%)	*VC* (%)	*NVC* (%)
Euler diagrams	97.2	81.6	85.7
Venn diagrams	88.3	69.3	59.8

subject-predicate order CA in conclusions). These data were subjected to angular transformation and a two-way ANOVA. There was a significant main effect for the difference between Euler and Venn diagrams, [F (1, 73) = 23.932. $p < 0.001$]. There was a significant main effect for the difference between single type syllogisms and multiple type syllogisms, [F (2, 146) = 35.144. $p < 0.001$]. There was a significant interaction effect for two factors, [F (2, 146) = 4.380. $p < 0.05$]. Post hoc tests showed the following results: (i) In the restricted Euler group, the performances of the valid syllogisms involving single type were better than those of the valid syllogisms involving multiple type [F (1, 88) = 5.251, $p < 0.05$], and than those of the NVC syllogisms involving multiple type [F (1, 88) = 3.707, $p < 0.05$]. (ii) The performances of the valid syllogisms involving single type in the restricted Euler group were better than those in the restricted Venn group at the reduced threshold of $p < 0.10$ [F (1, 219) = 3.808]. The performances of the valid syllogisms involving multiple type in the restricted Euler group were significantly better than those in the restricted Venn group, [$F(1, 219) = 6.228$, $p < 0.05$]. The performances of the NVC syllogisms (multiple type) in the restricted Euler group were significantly better than those in the restricted Venn group, [$F(1, 219) = 29.855$, $p < 0.001$].

We analyze a well-known interpretational bias in linguistic syllogistic reasoning: conversion errors and figural effects. Our results indicate that the two type of effects were blocked in both Euler and Venn diagrammatic groups.

In the 17 syllogisms having no valid conclusion, as Table 3 indicates, 58.8 % of the participants in the Linguistic group selected the illicit converted conclusion while

Table 3 The average error rates of conversion errors in the Linguistic, Euler and Venn group

	Linguistic (%)	Euler (%)	Venn (%)
Conversion error	58.8	20.5	36.0

Table 4 The average accuracy rates of syllogisms of EI1 and EI4 types in the Linguistic, Euler and Venn group

Figure	*1st & 2nd premises*	Linguistic (%)	Euler (%)	Venn (%)
EI1O	$no(B, A); some(C, B)$	62.2	77.8	48.5
EI4O	$no(A, B); some(B, C)$	35.6	68.3	42.6

the rate reduced to 20.5 % in the Euler group and 36.0 % in the Venn group. These data were also subjected to a one-way ANOVA. There was significant main effect, $[F(2, 173) = 26.864, p < 0.001]$. Multiple comparison tests yield the following results: (i) There was significant difference between the Linguistic group and the Euler group, $[F(1, 106) = 7.329, p < 0.001]$. (ii) There was significant difference between the Venn group and the Linguistic group, $[F(1, 111) = 4.441, p < 0.001]$. (iii) There was significant difference between the Euler group and the Venn group, $[F(1, 129) = 3.300, p < 0.005]$.

In fact, comparing EI1O and EI4O syllogisms, as Table 4 indicates, there was significant difference between EI1O$_{no(B, A); some(C, B) : some\text{-}not(C, A)}$ and EI4O$_{no(A, B); some(B, C) : some\text{-}not(C, A)}$ in the Linguistic group (62.2 % for EI1O and 35.6 % for EI4O) $[P = 0.0084$, Wilcoxon test]. In contrast, there was no significant difference between EI1O and EI4O in the Euler group (77.8 % for EI1O and 68.3 % for EI4O) $[P = 0.1578]$. Further, there was no significant difference between EI1O and EI4O in the Venn group (48.5 % for EI1O and 42.6 % for EI4O) $[P = 0.3464]$.

3.3 Discussion

The performance of syllogistic reasoning in the Euler group was significantly better than that in the Linguistic group. This shows that suitably formulated diagrammatic representations can be effective in supporting logical reasoning, in contrast to the prevailing claim in the literature that the use of circle diagrams is not helpful but rather sometimes harmful for logical reasoning (cf. Erickson 1974; Johnson-Laird 1983; Calvillo et al. 2006). Furthermore, the performance of the Euler group was significantly better than that of the Venn group; and this tendency held of all the three types of syllogisms, i.e., single and multiple type syllogisms with valid conclusion and syllogisms with no valid conclusion (which are always multiple-type). These results support our prediction of inferential efficacy, namely, the claim that the process of solving logical reasoning tasks can be replaced by manipulation of diagrams. Note that in our experiment, the participants in the Euler group received more instruction and 10 trials of practice (in diagram interpretation, rather than syllogism solving), whereas people in the Linguistic group did not. It might be claimed that this difference in training has had a major impact on their differences in the performance in syllogism solving. However, this potential objection can be avoided by comparing between the Euler group and the Venn group, since both groups received substantial instruction about the categorical sentences and underwent a similar number of practice trials. The observed difference between the performance of the Euler group and that of the Venn group would suggest that the results were not caused by practice effects alone; rather, a reasonable explanation is that Euler diagrams not only contributes to the correct interpretations of categorical sentences but also plays a substantial role in the inferential processes of syllogism solving.

Moreover, as for the Euler group, the performance in single-type syllogisms (i.e., those in which unification determines a unique diagram) was significantly better than that in multiple-type syllogisms (i.e., those in which unification gives rise to multiple cases). This confirms the prediction that solving the multiple-type syllogisms requires more efforts than solving the single-type syllogisms, which in turn provides partial evidence that processes of manipulating diagrams, or, more specifically, process of unifying premise diagrams, are involved in the overall processes of conducting syllogistic reasoning with diagrams. These findings could be naturally explained by positing the cognitive model of syllogistic reasoning supported by diagrams as we stated in Sect. 2. According to this model, Euler diagrams are effective in supporting syllogistic reasoning by virtue of the fact that they are effective ways of representing and reasoning about relational structures that are implicit in categorical sentences. All the results for the Euler group obtained in the experiments would support this conclusion.

The performance of the Venn group might be explained by the contribution of Venn diagrams to participants' interpretations of categorical sentences, while not playing a substantial role in reasoning processes themselves.[16] Hence, participants in the Venn group would have to rely on inferences based on abstract semantic information extractable from sentences and diagrams rather than concrete syntactic manipulations of diagrams. This explanation is in accord with the finding that errors caused by illicit conversion of categorical sentences in the Linguistic group were significantly reduced in the Venn group. In addition, a figural effect in the performance between EI1O and EI4O syllogisms was observed in the Linguistic group but not in the Venn group. As expected from the prediction of the interpretational efficacy (conversion effect and figural effects), these results may be explained by Venn diagrams aiding participants' interpretations, preventing certain well-known interpretational errors in syllogisms caused by the word order of categorical sentences.

4 Additional Analysis

Our Experiment provided some evidence to support the claim that the efficacy of Euler diagrams depends on the fact that they are effective ways of representing and reasoning

[16] In our Experiment, the participants in the Venn group were provided with diagrams consisting of *two* circles that corresponded to the premises of a given syllogism. However, we also tested a situation where participants are initially provided with Venn diagrams consisting of *three* circles, or "3-Venn diagrams", namely D_3 and D_4 of Fig. 12. With 3-Venn diagrams, participants could skip the first steps of adding new circles; the only step needed is to superpose the two premise diagrams. Thus, it may be predicted that 3-Venn diagrams are relatively easy to manipulate in syllogism solving, even for novices. In fact, in the experiments of Sato et al. (2010), we obtained results confirming this prediction.

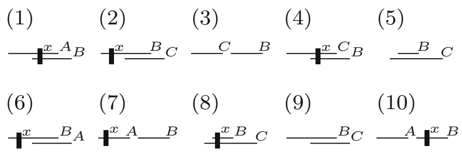


Fig. 22 Linear diagrams used in the pretest

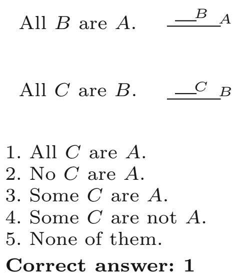


Fig. 23 An example of syllogistic reasoning task of the Linear group

about the relational information contained in syllogistic sentences (for example, set inclusion and set disjointness are effectively represented by the corresponding relationship between circles). If this claim is correct, it would be expected that any diagram that makes explicit the relational information encoded in a categorical sentence in a suitable way could be as effective in supporting syllogistic reasoning as Euler diagrams. Additional analysis was designed to confirm this prediction. For this purpose, we used a linear variant of Euler diagrams we used in the Experiment. We compared the performances of solving syllogisms using Linear diagrams with the performances of Linguistic group, Euler group, and Venn group.

4.1 Method

The additional analysis was conducted in the same manner as the Experiment; the only difference is that in syllogistic reasoning tasks, premise sentences were associated with corresponding linear diagrams, instead of Euler diagrams or Venn diagrams. Thirty-four undergraduates (mean age 22.72 ± 8.72 SD) in an elementary philosophy class participated. We called this group "the Linear group". Of 34 participants, we excluded 6 participants: those who left the last more than three questions unanswered (5 participants) and those who did not follow our instructions (1 participant). Figure 22 is a list of Linear diagrams used in the pretest. Figure 23 is an example of syllogistic reasoning task of the Linear group.

4.2 Results and Discussion

4.2.1 Pretest

The accuracy rate of each item in the pretest of linear diagrams, listed in Fig. 22. was (1) 92.9 %, (2) 85.7 %, (3) 96.0 %, (4) 89.3 %, (5) 75.0 %, (6) 86.0 %, (7) 85.7/53.6 % (46.0 %), (8) 82.1/78.6 % (61.0 %), (9) 82.1 %, and (10) 64.3 %, respectively. Seven participants scored less than 8 on the pretest (out of 12). In the following analysis, we exclude these 7 participants and refer to the other 21 participants as the restricted Linear group.

4.2.2 Syllogistic Reasoning Tasks

The average accuracy rate for the 31 syllogisms in the Linear group was 71.2 %. The results of each syllogistic type are shown in "Appendix 1". The Linear group data were compared with the Linguistic, Venn, and Euler groups in the Experiment using one-way ANOVA. The results revealed a significant main effect, F (3, 200) = 27.298, $p < 0.001$. Multiple comparison tests using Ryan's procedure yielded the following results: (i) The accuracy rate of reasoning tasks in the Linear group was significantly higher than in the Linguistic group: 46.7 % for the Linguistic group and 71.1 % for the Linear group [F (1, 71) = 4.527, $p < 0.001$]. (ii) The accuracy rate of reasoning tasks in the Linear group was higher than in the Venn group: 52.9 % for the Venn group and 71.2 % for the Linear group [F (1, 94) = 4.625, $p < 0.001$]. (iii) There was no significant difference between the accuracy rate for reasoning tasks in the Linear group and in the Euler group: 80.7 % for the Linear group and 77.8 % for the Euler group [F (1, 89) = 1.287, $p > 0.1$]. The accuracy rate for the restricted Linear group (excluding those who failed the pretest) was 80.9 %. These results showed that the performance of syllogistic reasoning in the Linear groups was significantly better than that in the Venn groups.

In the restricted Linear group, the average accuracy rate for valid syllogisms involving single type was 96.4 %, the accuracy rate for valid syllogisms involving multiple type was 76.6 %, and the accuracy rate for NVC syllogisms involving multiple type was 80.9 %. These data were subjected to a one-way ANOVA. There was a significant main effect, F (2, 62) = 6.670, $p < 0.001$. Multiple comparison tests conducted yielded the following results: performance for valid syllogisms involving single type were significantly better than for valid syllogisms involving multiple type [F (1, 20) = 3.472, $p < 0.005$], and for NVC syllogisms involving multiple type [F (1, 20) = 2.717, $p < 0.05$]. This showed that as for the restricted Linear group, the performance of solving single-type syllogisms was significantly better than the performance of solving multiple-type syllogisms, as in the case with Euler diagrams. In addition, performance with NVC syllogisms in the Linear group was significantly better than in the restricted Venn group: 80.9 % for the restricted Linear group and 59.5 % for the restricted Venn group [F (1, 50) = 3.639, $p < 0.05$]. Thus, the performances of NVC tasks were significantly better when linear diagrams were provided, as compared to Venn diagrams.

The additional analysis shows that linear diagrams function as effectively as Euler diagrams in syllogistic reasoning, and accordingly, that the effectiveness of external diagrams in syllogistic reasoning is not due to the particular shape of set diagrams. These findings provides further evidence to support the claim that the efficacy of external diagrams in syllogistic reasoning derives from the fact that they are concrete methods of representing and reasoning about relational information (i.e., inclusion and exclusion relations) encoded by categorical sentences in a way that is suitable for syntactic manipulation.

5 General Discussion

As we saw in Introduction, it has been claimed in the previous literature that diagrams are not necessarily effective in improving deductive reasoning but rather sometimes defective methods to support people's reasoning. Contrary to this widespread view, our experiments suggested that suitably formulated diagrams provide effective and concrete ways of solving syllogistic reasoning tasks. By "suitably formulated" diagrams, we mean diagrammatic representation systems in which there is a one-to-one correspondence between diagrams and sentences. In contrast to traditional Euler circles (Gergonne's diagrams), there is a direct correspondence between the Euler diagrams we used in our experiments and syllogistic sentences. More specifically, Euler diagrams and their linear variants represent the underlying semantic relations implicit in categorical sentences in terms of their spatial relations, i.e., inclusion and exclusion relations between points and circles. Accordingly, reasoning about set-theoretical relations can be replaced by reasoning about spatial relations between concrete objects. This makes it possible not only to prevent the misinterpretation of categorical sentences but also to replace reasoning about set-theoretical relations by reasoning about spatial relations between concrete objects. For diagrams to be effective in syllogistic reasoning, it is essential to mirror the relational information encoded by categorical sentences so as to be able to simulate reasoning about semantic information in terms of concrete manipulation of diagrams.

The efficacy of various forms of Euler diagrams has been investigated in the study of logic teaching methods (e.g., Monaghan and Stenning 1998; Dobson 1999, mentioned in Sect. 2.3.3). In these experimental studies, participants are typically provided with substantial (about an hourlong) training about manipulating ways of diagrams. By contrast, our study examined the question of whether diagrams can be useful for people who are not trained in rules or strategies of diagrammatic reasoning. This question is important because, in contrast to logical formulas in symbolic logic, Euler diagrams in general are considered more intuitive and effective for novices' reasoning, rather than for experts' or machine reasoning. Our experiments support the claim that a natural class of reasoning processes with Euler diagrams, such as processes of identifying objects and unifying two diagrams, can be conducted without explicit knowledge of the underlying rule or strategies. In such cases, diagrammatic inferences can be naturally triggered based on the correct understanding of the meaning of diagrams.

As we argued in Sect. 2.3.3, in addition, Euler diagrams can contribute to solving the NVC tasks by making available to users syntactic processes of unifying diagrams. The process of checking invalidity can be performed in terms of a unification process in a similar way to that of checking validity. Thus, the procedure of checking invalidity sketched here is remarkably distinguished from the standard procedure in model-theoretic semantics, according to which an inference is judged to be invalid if one can construct a counter-model in which all the premises are true but the conclusion is false. An interesting feature of the unification processes is that premise diagrams themselves impose a constraint on the possible ways of unification, so that by simply trying to unify the premise diagrams, the user could observe what relations hold between the objects in the resulting diagram. In this respect, our discussion here are consistent with the basic claim of the mental model theory (Johnson-Laird 1983; Johnson-Laird and Byrne 1991; for a formal specification, see Sugimoto and Sato 2015) that logically untrained people tend to make errors with inferences that require *multiple* models. If, as we argued, a process of unifying premise diagrams triggers a process of entertaining alternative possibilities, then it would increase the chances of finding alternative situations that are compatible with all the premises. Furthermore, diagrams can serve as a memory-aid in keeping track of alternative models (Bauer and Johnson-Laird 1993). Thus, in sum, diagrams that are externally given could help a user to find and keep track of the relevant alternative situations through a process of unification (see also Mineshima et al. 2014 for a comparison between mental models and Euler diagrams).

In logical reasoning tasks using diagrammatic representations, our analysis distinguishes inference processes from interpretation processes in logical reasoning tasks. To separate two main factors contributing to solving reasoning tasks has attracted attention in the recent interaction between logic and cognitive science. According to Stenning and van Lambalgen (2001, 2004) as mentioned in Sect. 2.1, reasoning *toward* an interpretation of premises and reasoning *from* a fixed interpretation of premises play a crucial role. Stenning and van Lambalgen (2004) fixed the interpretation of natural language conditionals through experimenters' interventions (tutorial dialogues) and then analyzed the subjects' responses of drawing conclusion in conditional reasoning (Wason's selection) tasks. Politzer and Mercier (2008) used syllogisms with singular premises such as *this X is Y*, instead of normal existential quantified premise such as *some X is Y*, so as to control the interpretation of quantified sentences. Our exploration of reasoning tasks with diagrams can be put in the same line. In our setting, diagrams that are externally given to reasoners control the intended meaning of premise sentences, thus contributing to a proper understanding of what is essential about the inferential processes in question.

The finding that Euler diagrams significantly improve people's ability of logical reasoning would shed a new light on the conception of reasoning that has been widely accepted within the dual process theory (cf. Evans 2003, 2008). According to the standard account, the ability to conduct logical reasoning belongs to System 2, a cognitive system concerned with slow, analytic, reflective, and effortful processes, rather than to System 1, which is concerned with quick, intuitive, automated, and effortless processes. The fact that people make errors in logical reasoning tasks is then accounted for by assuming that they tend to rely on System 1 when trying to solve a given reasoning tasks [cf. Alter et al. 2007; see also Stenning and van Lambalgen 2008 where particular "errors" were viewed in classical logic but not in non-monotonic logic (closed world assumption)]. Interestingly, our experimental results can be taken as challenging this binary dichotomy: the results suggest that while reasoning processes with Euler diagrams are analytic in that they contribute to conducting accurate and correct reasoning, they are at the same time intuitive and automatic in that unifying premise diagrams leads to selecting a correct conclusion (cf. the free ride property discusses in Sect. 2.3.2).[17] In this sense, reasoning with Euler diagrams is a hybrid process having both features of System 1 and System 2.

In the discussion made so far, we argued that reasoning with the standard types of quantifiers, i.e., existential and universal quantifiers, is visualized in terms of spatial relations such as inclusion and exclusion. It is left for future work to see whether this framework is extended to other types of quantifiers, including proportional quantifiers such as *most*, numerical (cardinal) quantifier such as *at least three*, and comparative quantifier such as *more A than B*. Such non-standard quantifiers are more or less frequently used in ordinary reasoning (see Pfeifer and Kleiter 2005; Leslie and Gelman 2012 for discussion). In addition, multiply quantified sentences such as *All A R some B*, where *R* is a transitive verb, lead to a further complexity. Logical and computational aspects of such a complexity of quantifiers have been investigated in a systematic way within generalized quantifier theory (Barwise and Cooper 1981; Westerståhl 1989) and cognitive studies of quantifiers in natural language (Szymanik and Zajenkowski 2010). In the literature of psychology of human deduction, however, reasoning with such extended types of quantifiers and extended inference forms has attracted less attention; indeed, most studies have tended to focus on a limited class of quantified inferences, and hence, the cognitive property of non-standard type of quantifiers has not been well understood (cf. Khemlani and Johnson-Laird 2012). On the side of the logical study of reasoning with diagrams, various diagrammatic systems have developed to handle reasoning with extended types of quantifiers (for a recent survey, see Rodgers 2014). A common strategy is to add some novel syntactic devices to existing systems of diagrammatic representations. A natural question then arises how far we can go in this direction: it is certainly true that those systems that are combined with extended syntactic devices tend to be away from cognitively effective systems, and hence, come close to symbolic systems (see the literature on the so-called spider diagrams; Howse et al. 2005). To what extent can a diagrammatic system be expressive enough to handle the generality of logical inferences as well as be specific and effective so that it is easy to use even for untrained reasoners? Further research in this direction may lead to a better understanding of diagrammatic reasoning with sentences involving various levels of complexity, and, for that matter, of the relationship between linguistic and diagrammatic reasoning in general.

[17] Further evidence that diagrammatic reasoning has features inherent to System 2 comes from the fact that it can be faster than linguistic or sentential reasoning. Several experiments examined the solution time for reasoning tasks with diagrams and reported that it was short than that for reasoning tasks without diagrams; see Bauer and Johnson-Laird (1993), Cheng (2004), Sato et al. (2015).

Acknowledgments Parts of this study were presented at the 6th Diagrams Conference (August, 2010) in Portland and the 7th Diagrams Conference (July, 2012) in Canterbury. This study was partially supported by Grant-in-Aid for JSPS Fellows (21·7357; 25·2291) to the first author. The authors would like to thank Mitsuhiro Okada, Atsushi Shimojima, Ryo Takemura, and Keith Stenning for the valuable advice and Philip Johnson-Laird and Malcolm Bauer for permission to use unpublished results.

Appendix 1: The Results of Each Syllogistic Type

The results of each syllogistic type are shown in Tables 5 and 6. Numbers indicate the percentage of total responses to each syllogism. Bold type refers to valid conclusion by the standard of predicate logic. For simplicity, we exclude the conclusions of the so-called "weak" syllogisms (syllogisms whose validity depends on the existential import of subject term) from valid answers.

Table 5 Response distributions for 31 syllogisms in the Linguistic, Venn, Euler, Linear groups (Bold type refers to valid conclusion)

Code & figure	*Premises 1st, 2nd*	Linguistic group (N = 45) *Conclusion**					Venn group (N = 68) *Conclusion**					Euler group (N = 63) *Conclusion**					Linear group (N = 28) *Conclusion**				
		A	E	I	O	N	A	E	I	O	N	A	E	I	O	N	A	E	I	O	N
AA2	$all(A, B); all(C, B)$	55.6	6.7	4.4	2.2	**31.1**	23.5	1.5	22.1	7.4	**45.6**	6.3	1.6	3.2	0.0	**85.7**	7.1	0.0	14.3	0.0	**78.6**
AA3	$all(B, A); all(B, C)$	60.0	0.0	26.7	0.0	**13.3**	29.4	7.4	23.5	4.4	**35.3**	4.8	1.6	20.6	0.0	**73.0**	14.3	3.6	32.1	0.0	**50.0**
AA4	$all(A, B); all(B, C)$	60.0	0.0	28.9	0.0	**11.1**	25.0	8.8	25.0	10.3	**29.4**	6.3	0.0	22.2	1.6	**69.8**	3.6	0.0	42.9	0.0	**53.6**
AI1	$all(B, A); some(C, B)$	2.2	2.2	**88.9**	2.2	4.4	4.4	4.4	**70.6**	13.2	7.4	0.0	0.0	**98.4**	1.6	0.0	0.0	0.0	**96.4**	0.0	3.6
AI2	$all(A, B); some(C, B)$	0.0	0.0	55.6	17.8	**26.7**	1.5	8.8	42.6	5.9	**41.2**	0.0	1.6	12.7	7.9	**74.6**	0.0	0.0	17.9	10.7	**71.4**
AI3	$all(B, A); some(B, C)$	4.4	2.2	**80.0**	6.7	6.7	2.9	7.4	**70.6**	7.4	11.8	0.0	0.0	**85.7**	0.0	14.3	3.6	0.0	**82.1**	0.0	10.7
AI4	$all(A, B); some(B, C)$	4.4	0.0	57.8	6.7	**31.1**	5.9	1.5	41.2	10.3	**38.7**	0.0	0.0	19.0	7.9	**73.0**	0.0	0.0	21.4	7.1	**71.4**
IA1	$some(B, A); all(C, B)$	2.2	2.2	60.0	8.9	**26.7**	4.4	1.5	48.5	10.3	**35.3**	0.0	0.0	19.0	4.8	**73.0**	3.6	3.6	21.4	3.6	**67.9**
IA2	$some(A, B); all(C, B)$	6.7	0.0	51.1	11.1	**31.1**	8.8	10.3	29.4	7.4	**42.6**	0.0	0.0	23.8	1.6	**74.6**	0.0	0.0	25.0	3.6	**71.4**
IA3	$some(B, A); all(B, C)$	0.0	2.2	**93.3**	0.0	4.4	0.0	1.5	**76.5**	5.9	16.2	0.0	0.0	**79.4**	0.0	20.6	0.0	3.6	**78.6**	0.0	17.9
IA4	$some(A, B); all(B, C)$	11.1	0.0	**73.3**	6.7	6.7	4.4	14.7	**52.9**	5.9	22.1	0.0	0.0	**74.6**	0.0	25.4	0.0	3.6	**82.1**	3.6	10.7
AE1	$all(B, A); no(C, B)$	0.0	64.4	0.0	6.7	**26.7**	4.4	30.9	8.8	11.8	**42.6**	0.0	12.7	0.0	6.3	**79.4**	3.6	17.9	0.0	10.7	**67.9**
AE2	$all(A, B); no(C, B)$	2.2	**93.3**	0.0	2.2	2.2	0.0	**67.6**	2.9	17.6	11.8	0.0	**98.4**	0.0	0.0	1.6	0.0	**92.9**	3.6	3.6	0.0
AE3	$all(B, A); no(B, C)$	0.0	64.4	2.2	11.1	**22.2**	2.9	48.5	0.0	16.2	**29.4**	0.0	23.8	0.0	4.8	**71.4**	0.0	14.3	3.6	14.3	**67.9**
AE4	$all(A, B); no(B, C)$	0.0	**73.3**	4.4	6.7	15.6	1.5	**66.2**	10.3	8.8	10.7	0.0	**95.2**	0.0	1.6	3.2	0.0	**89.3**	0.0	7.1	3.6
EA1	$no(B, A); all(C, B)$	0.0	**91.1**	0.0	2.2	2.2	4.4	**79.4**	4.4	5.9	2.9	0.0	**96.8**	0.0	1.6	1.6	0.0	**89.3**	3.6	3.6	3.6
EA2	$no(A, B); all(C, B)$	2.2	**88.9**	0.0	4.4	4.4	5.9	**72.1**	11.8	8.8	1.5	0.0	**100.0**	0.0	0.0	0.0	0.0	**96.4**	0.0	3.6	0.0
EA3	$no(B, A); all(B, C)$	0.0	62.2	0.0	20.0	**17.8**	2.9	48.5	17.6	7.4	**22.1**	0.0	20.6	0.0	9.5	**69.8**	0.0	17.9	0.0	7.1	**75.0**
EA4	$no(A, B); all(B, C)$	2.2	60.0	2.2	13.3	**22.2**	0.0	44.1	7.4	20.6	**26.5**	0.0	15.9	0.0	11.1	**73.0**	3.6	21.4	0.0	10.7	**64.3**
AO1	$all(B, A); some\text{-}not(C, B)$	0.0	4.4	8.9	66.7	**17.8**	0.0	16.2	14.7	47.1	**22.1**	0.0	3.2	1.6	33.3	**61.9**	0.0	3.6	10.7	21.4	**64.3**
AO2	$all(A, B); some\text{-}not(C, B)$	0.0	4.4	4.4	**75.6**	15.6	1.5	13.2	8.8	54.4	20.6	0.0	0.0	4.8	**85.7**	9.5	0.0	3.6	3.6	**75.0**	17.9
AO3	$all(B, A); some\text{-}not(B, C)$	0.0	2.2	17.8	53.3	**26.7**	1.5	10.3	19.1	30.9	**38.2**	0.0	1.6	15.9	9.5	**68.3**	0.0	7.1	10.7	21.4	**60.7**

Table 5 continued

Code & figure	*Premises 1st, 2nd*	Linguistic group (N = 45) *Conclusion**					Venn group (N = 68) *Conclusion**					Euler group (N = 63) *Conclusion**					Linear group (N = 28) *Conclusion**				
		A	E	I	O	N	A	E	I	O	N	A	E	I	O	N	A	E	I	O	N
AO4	*all(A, B); some-not(B, C)*	0.0	4.4	11.1	55.6	**28.9**	1.5	10.3	11.8	26.5	**48.5**	0.0	3.2	4.8	12.7	**79.4**	0.0	7.1	3.6	25.0	**64.3**
OA1	*some-not(B, A); all(C, B)*	2.2	2.2	4.4	66.7	**24.4**	0.0	5.9	10.3	47.1	**36.8**	0.0	4.8	1.6	25.4	**68.3**	0.0	7.1	3.6	25.0	**64.3**
OA2	*some-not(A, B); all(C, B)*	2.2	2.2	4.4	64.4	**26.7**	4.4	13.2	13.2	27.9	**41.2**	0.0	9.5	4.8	19.0	**66.7**	0.0	14.3	3.6	17.9	**64.3**
OA3	*some-not(B, A); all(B, C)*	0.0	4.4	11.1	**80.0**	4.4	0.0	14.7	10.3	**52.9**	20.6	0.0	4.8	7.9	**66.7**	20.6	0.0	3.6	10.7	**42.9**	42.9
OA4	*some-not(A, B); all(B, C)*	0.0	11.1	20.0	42.2	**26.7**	4.4	13.2	19.1	21.6	**39.7**	0.0	4.8	11.1	9.5	**74.6**	0.0	10.7	7.1	10.7	**71.4**
EI1	*no(B, A); some(C, B)*	0.0	22.2	2.2	**62.2**	13.3	0.0	27.9	7.4	**48.5**	16.2	0.0	12.7	0.0	**77.8**	9.5	0.0	14.3	0.0	**71.4**	14.3
EI2	*no(A, B); some(C, B)*	0.0	26.7	4.4	**46.7**	22.2	2.9	11.8	11.8	**55.9**	16.2	0.0	12.7	1.6	**74.6**	11.1	0.0	14.3	3.6	**67.9**	14.3
EI3	*no(B, A); some(B, C)*	0.0	20.0	2.2	**53.3**	17.8	0.0	25.0	7.4	**55.9**	11.8	0.0	12.7	0.0	**73.0**	14.3	3.6	14.3	0.0	**50.0**	32.1
EI4	*no(A, B); some(B, C)*	0.0	28.9	6.7	**35.6**	28.9	0.0	32.4	8.8	**42.6**	16.2	0.0	11.1	0.0	**68.3**	19.0	0.0	14.3	0.0	**64.3**	21.4

* *Conclusion*, A: *all(C, A)*, E: *no(C, A)*, I: *some(C, A)*, O: *some-not(C, A)*, N: *no-valid*

Table 6 Response distributions for 31 syllogisms in the restricted Venn, restricted Euler, restricted Linear groups

Code & figure	*Premises 1st, 2nd*	Restricted Venn group (N = 30) *Conclusion**					Restricted Euler group (N = 45) *Conclusion**					Restricted Linear group (N = 21) *Conclusion**				
		A	E	I	O	N	A	E	I	O	N	A	E	I	O	N
AA2	*all*(*A*, *B*); *all*(*C*, *B*)	23.3	0.0	3.3	0.0	**73.3**	4.4	0.0	0.0	0.0	**95.6**	4.8	0.0	4.8	0.0	**90.5**
AA3	*all*(*B*, *A*); *all*(*B*, *C*)	16.7	0.0	16.7	0.0	**63.3**	0.0	0.0	13.3	0.0	**86.7**	4.8	0.0	33.3	0.0	**61.9**
AA4	*all*(*A*, *B*); *all*(*B*, *C*)	23.3	3.3	16.7	3.3	**53.3**	8.9	0.0	11.1	2.2	**77.8**	0.0	0.0	33.3	0.0	**66.7**
AI1	*all*(*B*, *A*); *some*(*C*, *B*)	0.0	0.0	**90.0**	0.0	6.7	0.0	0.0	**100.0**	0.0	0.0	0.0	0.0	**95.2**	0.0	4.8
AI2	*all*(*A*, *B*); *some*(*C*, *B*)	0.0	3.3	30.0	0.0	**66.7**	0.0	0.0	6.7	2.2	**88.9**	0.0	0.0	9.5	4.8	**85.7**
AI3	*all*(*B*, *A*); *some*(*B*, *C*)	0.0	0.0	**80.0**	6.7	13.3	0.0	0.0	**84.4**	0.0	15.6	0.0	0.0	**81.0**	0.0	14.3
AI4	*all*(*A*, *B*); *some*(*B*, *C*)	0.0	0.0	33.3	0.0	**66.7**	0.0	0.0	15.6	0.0	**84.4**	0.0	0.0	14.3	0.0	**85.7**
IA1	*some*(*B*, *A*); *all*(*C*, *B*)	3.3	0.0	33.3	0.0	**63.3**	0.0	0.0	8.9	2.2	**84.4**	0.0	0.0	19.0	0.0	**81.0**
IA2	*some*(*A*, *B*); *all*(*C*, *B*)	3.3	3.3	23.3	0.0	**70.0**	0.0	0.0	13.3	2.2	**84.4**	0.0	0.0	19.0	0.0	**81.0**
IA3	*some*(*B*, *A*); *all*(*B*, *C*)	0.0	0.0	**66.7**	0.0	33.3	0.0	0.0	**75.6**	0.0	24.4	0.0	0.0	**81.0**	0.0	19.0
IA4	*some*(*A*, *B*); *all*(*B*, *C*)	0.0	3.3	**50.0**	3.3	43.3	0.0	0.0	**68.9**	0.0	31.1	0.0	0.0	**85.7**	0.0	14.3
AE1	*all*(*B*, *A*); *no*(*C*, *B*)	0.0	33.3	3.3	3.3	**60.0**	0.0	6.7	0.0	4.4	**88.9**	0.0	14.3	0.0	4.8	**81.0**
AE2	*all*(*A*, *B*); *no*(*C*, *B*)	0.0	**86.7**	0.0	6.7	6.7	0.0	**97.8**	0.0	0.0	2.2	0.0	**100.0**	0.0	0.0	0.0
AE3	*all*(*B*, *A*); *no*(*B*, *C*)	3.3	36.7	0.0	10.0	**46.7**	0.0	15.6	0.0	4.4	**80.0**	0.0	9.5	4.8	4.8	**81.0**
AE4	*all*(*A*, *B*); *no*(*B*, *C*)	0.0	**80.0**	0.0	6.7	10.0	0.0	**95.6**	0.0	2.2	2.2	0.0	**95.2**	0.0	4.8	0.0
EA1	*no*(*B*, *A*); *all*(*C*, *B*)	0.0	**96.7**	0.0	0.0	3.3	0.0	**97.8**	0.0	2.2	0.0	0.0	**90.5**	0.0	4.8	4.8
EA2	*no*(*A*, *B*); *all*(*C*, *B*)	3.3	**90.0**	3.3	3.3	0.0	0.0	**100.0**	0.0	0.0	0.0	0.0	**100.0**	0.0	0.0	0.0
EA3	*no*(*B*, *A*); *all*(*B*, *C*)	0.0	53.3	0.0	10.0	**36.7**	0.0	11.1	0.0	4.4	**84.4**	0.0	9.5	0.0	4.8	**85.7**
EA4	*no*(*A*, *B*); *all*(*B*, *C*)	0.0	43.3	3.3	10.0	**43.3**	0.0	6.7	0.0	11.1	**82.2**	0.0	14.3	0.0	4.8	**81.0**
AO1	*all*(*B*, *A*); *some-not*(*C*, *B*)	0.0	0.0	10.0	60.0	**30.0**	0.0	4.4	0.0	20.0	**75.6**	0.0	0.0	9.5	14.3	**76.2**
AO2	*all*(*A*, *B*); *some-not*(*C*, *B*)	0.0	3.3	6.7	**66.7**	23.3	0.0	0.0	2.2	**91.1**	6.7	0.0	0.0	0.0	**85.7**	14.3

Table 6 continued

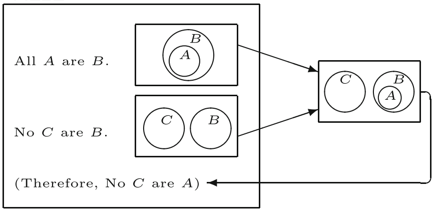
Code & figure	*Premises 1st, 2nd*	Restricted Venn group *Conclusion**				(N = 30)	Restricted Euler group *Conclusion**				(N = 45)	Restricted Linear group *Conclusion**				(N = 21)
		A	E	I	O	N	A	E	I	O	N	A	E	I	O	N
AO3	$all(B, A)$; $some\text{-}not(B, C)$	0.0	3.3	6.7	20.0	**70.0**	0.0	0.0	8.9	6.7	**77.8**	0.0	0.0	9.5	19.0	**71.4**
AO4	$all(A, B)$; $some\text{-}not(B, C)$	0.0	3.3	0.0	13.3	**83.3**	0.0	2.2	0.0	6.7	**91.1**	0.0	4.8	0.0	14.3	**81.0**
OA1	$some\text{-}not(B, A)$; $all(C, B)$	0.0	0.0	6.7	30.0	**63.3**	0.0	4.8	0.0	13.3	**82.2**	0.0	4.8	0.0	14.3	**81.0**
OA2	$some\text{-}not(A, B)$; $all(C, B)$	0.0	6.7	3.3	26.7	**63.3**	0.0	6.7	2.2	8.9	**82.2**	0.0	4.8	4.8	9.5	**81.0**
OA3	$some\text{-}not(B, A)$; $all(B, C)$	0.0	0.0	3.3	**60.0**	36.7	0.0	4.4	4.4	**66.7**	24.4	0.0	0.0	4.8	**47.6**	47.6
OA4	$some\text{-}not(A, B)$; $all(B, C)$	3.3	3.3	6.7	16.7	**70.0**	0.0	2.2	4.4	2.2	**91.1**	0.0	0.0	9.5	9.5	**81.0**
EI1	$no(B, A)$; $some(C, B)$	0.0	8.9	0.0	**70.0**	6.7	0.0	8.9	0.0	**84.4**	6.7	0.0	0.0	0.0	**85.7**	14.3
EI2	$no(A, B)$; $some(C, B)$	3.3	10.0	0.0	**73.3**	10.0	0.0	8.9	0.0	**84.4**	6.7	0.0	9.5	0.0	**76.2**	14.3
EI3	$no(B, A)$; $some(B, C)$	0.0	20.0	0.0	**63.3**	13.3	0.0	4.4	0.0	**84.4**	11.1	0.0	14.3	0.0	**57.1**	28.6
EI4	$no(A, B)$; $some(B, C)$	0.0	10.0	0.0	**73.3**	16.7	0.0	6.7	0.0	**75.6**	15.6	0.0	4.8	0.0	**71.4**	23.8

* *Conclusion*, A: $all(C, A)$, E: $no(C, A)$, I: $some(C, A)$, O: $some\text{-}not(C, A)$, N: *no-valid*

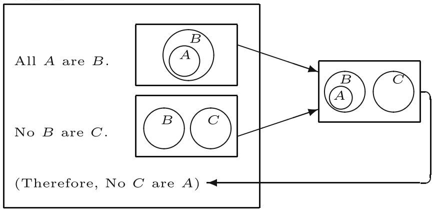
Appendix 2: Solving Processes Using Euler Diagrams

Single Type

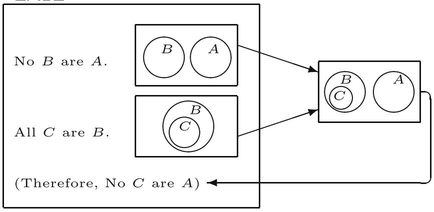
AE2E



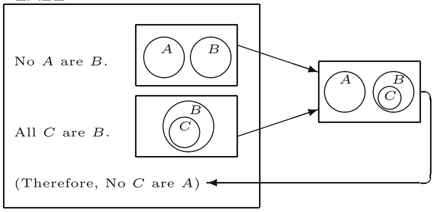
AE4E



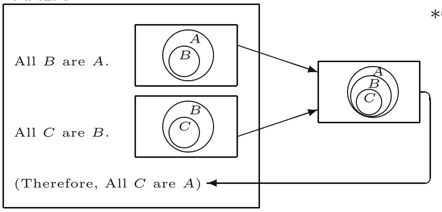
EA1E



EA2E



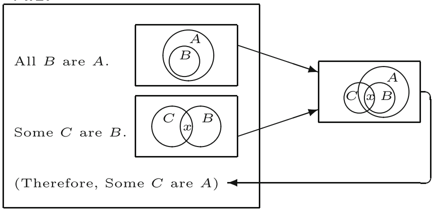
AA1A*



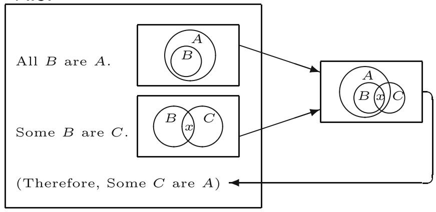
*This type was used as an example of task.

Multiple Type (Valid)

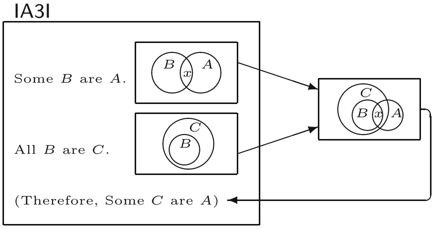
AI1I



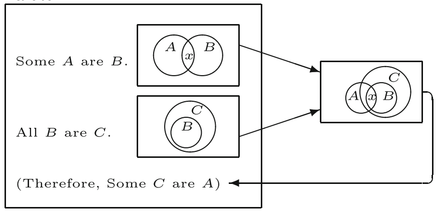
AI3I



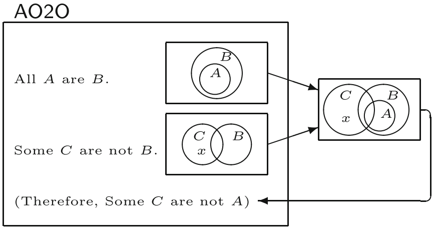
IA3I



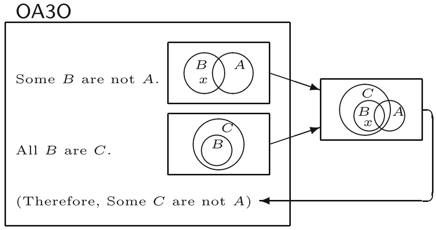
IA4I



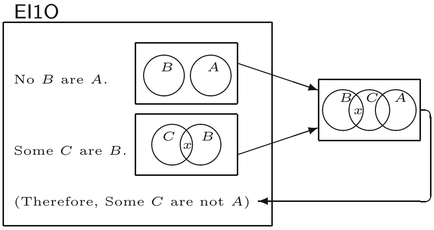
AO2O



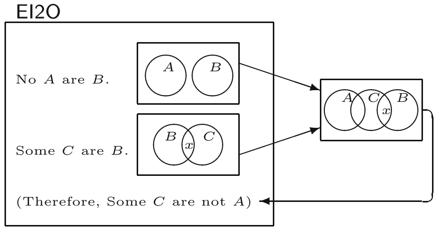
OA3O



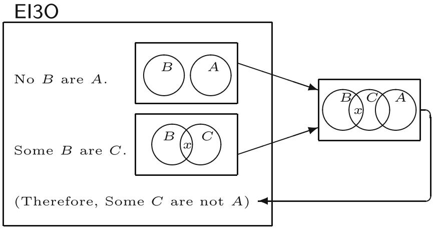
EI1O



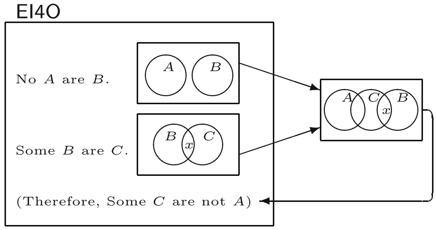
EI2O



EI3O

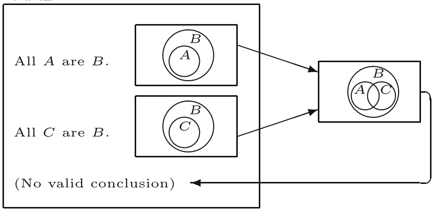


EI4O

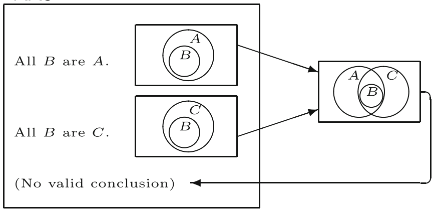


Multiple Type (Invalid)

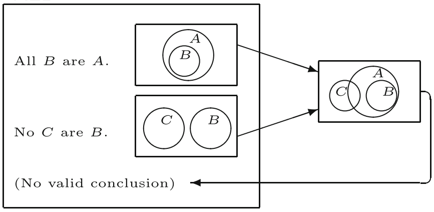
AA2



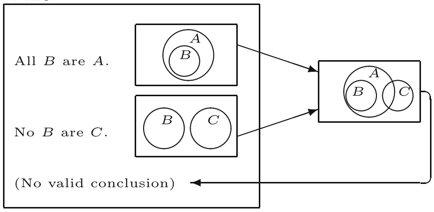
AA3



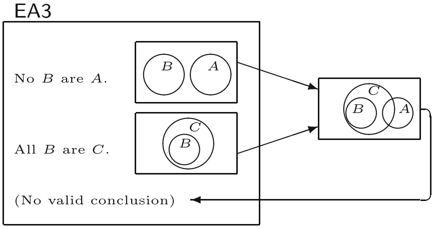
AE1



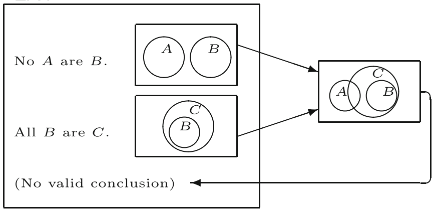
AE3



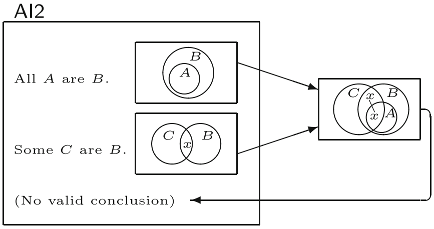
EA3



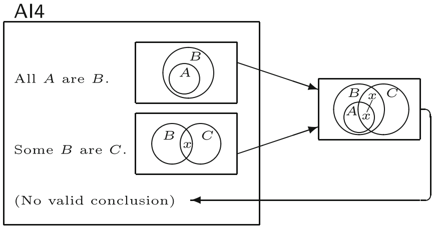
EA4



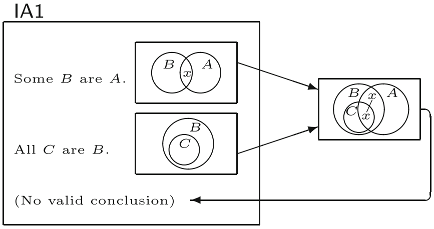
AI2



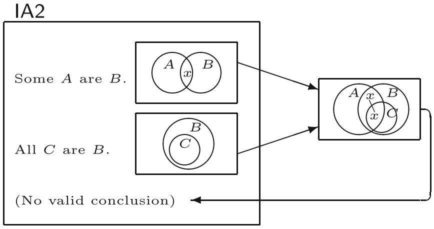
AI4

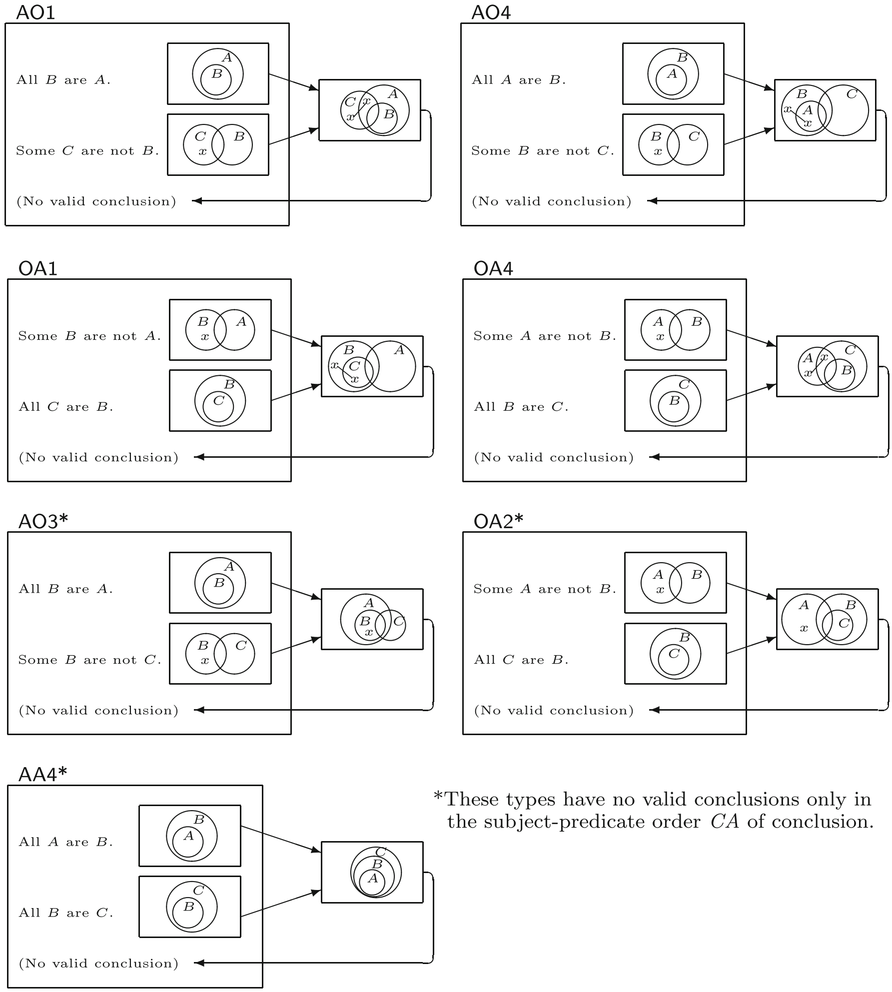


IA1



IA2





References

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*, 569–576.

Baron, M. E. (1969). A note on the historical development of logic diagrams: Leibniz, Euler and Venn. *The Mathematical Gazette*, *53*, 113–125.

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, *4*, 159–219.

Barwise, J., & Etchemendy, J. (1991/1996). Visual information and valid reasoning. In W. Zimmerman & S. Cunningham (Eds.), *Visualization in teaching and learning mathematics*, *MAA Notes, Number 19* (pp. 9–24). Reprinted in G. Allwein & J. Barwise (Eds). (1996). *Logical reasoning with diagrams* (pp. 3–26). New York: Oxford University Press.

Barwise, J., & Etchemendy, J. (1994). *Hyperproof, CSLI Lecture Notes, No. 42*. Stanford, CA: CSLI Publications.

Barwise, J., & Hammer, E. (1996). Diagrams and the concept of logical system. In G. Allwein & J. Barwise (Eds.), *Logical reasoning with diagrams* (pp. 49–78). New York: Oxford University Press.

Bauer, M., & Johnson-Laird, P. N. (1993). How diagrams can improve reasoning. *Psychological Science*, *4*, 372–378.

Benoy, F., & Rodgers, P. (2007). Evaluating the comprehension of Euler diagrams. In E. Banissi, et al. (Eds.), *Proceedings of information visualization 2007* (pp. 771–778). Los Alamitos, CA: IEEE Computer Society.

Blackburn, P., & Bos, J. (2005). *Representation and inference for natural language: A first course in computational semantics*. Stanford, CA: CSLI Publications.

Boolos, G. (1984). On 'syllogistic inference'. *Cognition*, *17*, 181–182.

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

Calvillo, D. P., DeLeeuw, K., & Revlin, R. (2006). Deduction with Euler circles: Diagrams that hurt. In D. Baker-Plummer, R. Cox, & N. Swoboda (Eds.), *Proceedings of Diagrams 2006, LNAI 4045* (pp. 199–203). Berlin, Heidelberg: Springer.

Camp, E. (2007). Thinking with maps. *Philosophical Perspectives*, *21*, 145–182.

Chapman, L., & Chapman, J. (1959). Atmosphere effect re-examined. *Journal of Experimental Psychology*, *58*, 220–226.

Chapman, P., Stapleton, G., Rodgers, P., Micallef, L., & Blake, A. (2014). Visualizing sets: An empirical comparison of diagram types. In T. Dwyer, H. Purchase, & A. Delaney (Eds.), *Proceedings of Diagrams 2014, LNAI 8578* (pp. 146–160). Berlin, Heidelberg: Springer.

Cheng, P. C.-H. (2004). Why diagrams are (sometimes) six times easier than words: Benefits beyond locational indexing. In A. F. Blackwell, K. Marriott, & A. Shimojima (Eds.), *Proceedings of Diagrams 2004, LNAI 2980* (pp. 242–260). Berlin, Heidelberg: Springer.

Deloache, J. S., Sugarman, S., & Brown, A. L. (1985). The development of error correction strategies in young children's manipulative play. *Child Development*, *56*, 928–939.

Dickstein, L. S. (1978). The effect of figure on syllogistic reasoning. *Memory and Cognition*, *6*, 76–83.

Dickstein, L. S. (1981). The meaning of conversion in syllogistic reasoning. *Bulletin of the Psychonomic Society*, *18*, 135–138.

Dobson, M. (1999). Information enforcement and learning with interactive graphical systems. *Learning and Instruction*, *9*, 365–390.

Doumas, L. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review*, *115*, 1–43.

Edwards, A. W. F. (2004). *Cogwheels of the mind: The story of venn diagrams*. Baltimore, MD: The Johns Hopkins University Press.

Englebretsen, G. (1992). Linear diagrams for syllogisms (with relationals). *Notre Dame Journal of Formal Logic*, *33*, 37–69.

Erickson, J. R. (1974). A set analysis theory of behavior in formal syllogistic reasoning tasks. In R. Solso (Ed.), *Loyola symposium on cognition* (Vol. 2, pp. 305–329). Hillsdale, NJ: Erlbaum.

Euler, L. (1768). *Lettres à une Princesse d'Allemagne sur Divers Sujets de Physique et de Philosophie*. Saint-Pétersbourg: De l'Académie des Sciences.

Evans, J St B T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*, 454–459.

Evans, J. S. B. T. (2008). Dual processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278.

Evans, J. S. B. T., Newstead, S. E., & Byrne, R. J. (1993). *Human reasoning: The psychology of deduction*. Hove: Erlbaum.

Fish, A., Khazaei, B., & Roast, C. (2011). User-comprehension of Euler diagrams. *Journal of Visual Languages and Computing*, *22*, 340–354.

Ford, M. (1994). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition*, *54*, 1–71.

Gattis, M. (2005). Inferencing from spatial information. *Spatial Cognition and Computation*, *5*, 119–137.

Gergonne, J. D. (1817). Essai de dialectique rationelle. *Annuales de Mathematiques Pures et Appliqukes*, *7*, 189–228.

Geis, M. L., & Zwicky, A. M. (1971). On invited inferences. *Linguistic Inquiry*, *2*, 561–566.

Geurts, B. (2003). Reasoning with quantifiers. *Cognition*, *86*, 223–251.

Geurts, B. (2007). Existential import. In I. Comorovski & K. von Heusinger (Eds.), *Existence: Semantics and syntax* (pp. 253–271). Netherlands: Springer.

Glasgow, J., Narayanan, N. H., & Chandrasekaran, B. (Eds.). (1995). *Diagrammatic reasoning: Cognitive & computational perspectives*. Cambridge, MA: AAAI Press/MIT Press.

Greenfield, P. M., Nelson, K., & Saltzman, E. (1972). The development of rulebound strategies for manipulating seriated cups: A parallel between action and grammar. *Cognitive Psychology*, *3*, 291–310.

Gurr, C. A., Lee, J., & Stenning, K. (1998). Theories of diagrammatic reasoning: Distinguishing component problems. *Minds and Machines*, *8*, 533–557.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*, 803–831.

Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences*, *14*, 497–505.

Hammer, E., & Shin, S. (1998). Euler's visual logic. *History and Philosophy of Logic*, *19*, 1–29.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, *8*, 280–285.

Heiser, J., & Tversky, B. (2006). Arrows in comprehending and producing mechanical diagrams. *Cognitive Science*, *30*, 581–592.

Hertzum, M., & Frøkjær, E. (1996). Browsing and querying in online documentation: A study of user interfaces and the interaction process. *ACM Transactions on Computer–Human Interaction*, *3*, 136–161.

Hodes, H. T. (1984). Logicism and the ontological commitments of arithmetic. *Journal of Philosophy*, *81*, 123–49.

Horn, L. R. (2000). From if to iff: Conditional perfection as pragmatic strengthening. *Journal of Pragmatics*, *32*, 289–326.

Howse, J., Molina, F., Shin, S.-J., & Taylor, J. (2002). On diagram tokens and types. In M. Hegarty, B. Meyer, & N. H. Narayanan (Eds.), *Proceedings of Diagrams 2002, LNAI 2317* (pp. 146–160). Berlin, Heidelberg: Springer.

Howse, J., Stapleton, G., & Taylor, J. (2005). Spider diagrams. *LMS Journal of Computation and Mathematics*, *8*, 145–194.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago, IL: The University of Chicago Press.

Johnson-Laird, P. N. (1983). *Mental models: Towards cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N., & Byrne, R. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.

Johnson-Pynn, J., Fragaszy, D. M., Hirsh, E. M., Brakke, K. E., & Greenfield, P. M. (1999). Strategies used to combine seriated cups by chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), and capuchins (*Cebus apella*). *Journal of Comparative Psychology*, *113*, 137–148.

Jones, S., McInnes, S., & Staveley, M. S. (1999). A graphical user interface for Boolean query specification. *International Journal of Digital Library*, *2*, 207–223.

Kneale, W., & Kneale, M. (1962). *The development of logic*. Oxford: Oxford University Press.

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

Lambert, J. H. (1764). *Neues Organon oder Gedanken über die Erforschung und Bezeichnung des Wahren und dessen Unterscheidung vom Irrtum und Schein*. Leipzig: Johann Wendler.

Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York, NY: Basic Books.

Larkin, J., & Simon, H. (1987). Why a diagram is (sometimes) worth 10,000 words. *Cognitive Science*, *11*, 65–99.

Leibniz, G.W. (1903/1988). *Opuscules et fragments inèdits: extraits des manuscrits de la bibliothèque royale de Hanovre*. L. Couturat (ed). Hildesheim: Georg Olms Verlag.

Lemon, O., & Pratt, I. (1998). On the insufficiency of linear diagrams for syllogisms. *Notre Dame Journal of Formal Logic*, *39*, 573–580.

Leslie, S. J., & Gelman, S. A. (2012). Quantified statements are recalled as generics: Evidence from preschool children and adults. *Cognitive Psychology*, *64*, 186–214.

Lull, R. (1617). *Opera Ea Quae Ad Adinventam Ab Ipso Artem Universalem, Scientiarum Artiumque Omnium Breui Compendio, Firmaq́; Memoria Apprehendendarum, Locupletiffimaq́; Vel Oratione Ex Tempore Pertractandarum, Pertinent*. Strassburg: Lazari Zetzneri.

MacFarlane, J. (2002). Frege, Kant, and the logic in logicism. *The Philosophical Review*, *111*, 25–65.

Meilinger, T., & Knauff, M. (2008). Ask for directions or use a map: A field experiment on spatial orientation and wayfinding in an urban environment. *Journal of Spatial Science*, *53*, 13–23.

Michard, A. (1982). Graphical presentation of boolean expressions in a database query language: Design notes and an ergonomic evaluation. *Behaviour and Information Technology*, *1*, 279–288.

Mineshima, K., Okada, M., Sato, Y., & Takemura, R. (2008). Diagrammatic reasoning system with Euler circles: Theory and experiment design. In G. Stapleton, J. Howse, & J. Lee (Eds.), *Proceedings of Diagrams 2008, LNAI 5223* (pp. 188–205). Berlin Heidelberg: Springer.

Mineshima, K., Okada, M., & Takemura, R. (2012a). A diagrammatic reasoning system with Euler circles. *Journal of Logic, Language and Information*, *21*, 365–391.

Mineshima, K., Okada, M., & Takemura, R. (2012b). A generalized syllogistic inference system based on inclusion and exclusion relations. *Studia Logica*, *100*, 753–785.

Mineshima, K., Sato, Y., Takemura, R., & Okada, M. (2014). Towards explaining the cognitive efficacy of Euler diagrams in syllogistic reasoning: A relational perspective. *Journal of Visual Languages and Computing*, *25*, 156–169.

Monaghan, P., & Stenning, K. (1998). Effects of representational modality and thinking style on learning to solve reasoning problems. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th annual conference of the cognitive science society* (pp. 716–721). Mahwah, NJ: Erlbaum.

Myers, K. L., & Konolige, K. (1994). Reasoning with analogical representations. In G. Lakemeyer & B. Nebel (Eds.), *Foundations of knowledge representation and reasoning, LNAI 810* (pp. 229–249). Berlin, Heidelberg: Springer.

Newstead, S. E., & Griggs, R. (1983). Drawing inferences from quantified statements: A study of the square of opposition. *Journal of Verbal Learning and Verbal Behavior*, *22*, 535–546.

Palmer, S. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 259–303). New Jersey: Lawrence Elbaum.

Peirce, C.S. (1897/1933). *Collected papers of Charles Sanders Peirce IV*. In C. Hartshorne & P. Weiss (Eds.), Cambridge, MA: Belknap Press of Harvard University Press.

Pfeifer, N., & Kleiter, G. D. (2005). Towards a mental probability logic. *Psychologica Belgica*, *45*, 71–100.

Politzer, G., van der Henst, J.-B., Luche, C. D., & Noveck, I. A. (2006). The interpretation of classically quantified sentences: A set-theoretic approach. *Cognitive Science*, *30*, 691–723.

Politzer, G., & Mercier, H. (2008). Solving categorical syllogisms with singular premises. *Thinking and Reasoning*, *14*, 414–453.

Pratt, I. (1993). Map semantics. In A. U. Frank & I. Campari (Eds.), *Spatial information theory: A theoretical basis for GIS, LNCS 716* (pp. 77–91). Berlin, Heidelberg: Springer.

Purchase, H. C. (2014). Twelve years of diagrams research. *Journal of Visual Languages and Computing*, *25*, 57–75.

Rescorla, M. (2009). Predication and cartographic representation. *Synthese*, *169*, 175–200.

Revlis, R. (1975). Two models of syllogistic reasoning: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior*, *14*, 180–195.

Rizzo, A., & Palmonari, M. (2005). The mediating role of artifacts in deductive reasoning. In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the 27th annual conference of the cognitive science society* (pp. 1862–1867). Austin, TX: Cognitive Science Society.

Roberts, M. J., Newstead, S. E., & Griggs, R. A. (2001). Quanntifier interpretation and syllogistic reasoning. *Thinking and Reasoning*, *7*, 173–204.

Rodgers, P. (2014). A survey of Euler diagrams. *Journal of Visual Languages and Computing*, *25*, 134–155.

Sato, Y., Mineshima, K., & Takemura, R. (2010). Constructing internal diagrammatic proofs from external logic diagrams. In R. Catrambone & S. Ohlsson (Eds.), *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 2668–2673). Austin, TX: Cognitive Science Society.

Sato, Y., Mineshima, K., & Takemura, R. (2011). Interpreting logic diagrams: A comparison of two formulations of diagrammatic representations. In C. Hoelscher, T. F. Shipley & L. Carlson (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 2182–2187). Austin, TX: Cognitive Science Society.

Sato, Y., Masuda, S., Someya, Y., Tsujii, T., & Watanabe, S. (2015). An fMRI analysis of the efficacy of Euler diagrams in logical reasoning. In *Proceedings of 2015 IEEE symposium on visual languages and human-centric computing* (9 pp). Los Alamitos, CA: IEEE Computer Society Press.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human–Computer Studies*, *45*, 185–213.

Shimojima, A. (1996). *On the efficacy of representation*. PhD thesis, Indiana University.

Shimojima, A., & Katagiri, Y. (2013). An eye-tracking study of exploitations of spatial constraints in diagrammatic reasoning. *Cognitive Science*, *37*, 211–254.

Shin, S.-J. (1994). *The logical status of diagrams*. New York: Cambridge University Press.

Stapleton, G. (2005). A survey of reasoning systems based on Euler diagrams. *Proceedings of Euler diagrams 2004, ENTCS 134* (pp. 127–151). Amsterdam: Elsevier.

Stenning, K. (2002). *Seeing reason: Image and language in learning to think*. Oxford: Oxford University Press.

Stenning, K., Cox, R., & Oberlander, J. (1995). Contrasting the cognitive effects of graphical and sentential logic teaching: Reasoning, representation and individual differences. *Language and Cognitive Processes*, *10*, 333–354.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning. *Cognitive Science*, *19*, 97–140.

Stenning, K., & Lemon, O. (2001). Aligning logical and psychological perspectives on diagrammatic reasoning. *Artificial Intelligence Review*, *15*, 29–62.

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.

Stenning, K., & van Lambalgen, M. (2004). A little logic goes a long way: Basing experiment on semantic theory in the cognitive science of conditional reasoning. *Cognitive Science*, *28*, 481–529.

Stenning, K., & van Lambalgen, M. (2008). *Human reasoning and cognitive science*. Cambridge, MA: MIT Press.

Sugimoto, Y., & Sato, Y. (2015). A specification-aware modeling of syllogistic reasoning with mental models. In N. A. Taatgen, M. K. van Vugt, J. P. Borst & K. Mehlhorn (Eds.), *Proceedings of the 13th international conference on cognitive modeling* (pp. 31–36). Groningen, The Netherlands: University of Groningen.

Szymanik, J., & Zajenkowski, M. (2010). Comprehension of simple quantifiers: Empirical evaluation of a computational model. *Cognitive Science*, *34*, 521–532.

Takemura, R. (2013). Proof theory for reasoning with Euler diagrams: A logic translation and normalization. *Studia Logica*, *101*, 157–191.

Venn, J. (1881). *Symbolic logic*. London: Macmillan.

Westerståhl, D. (1989). Quantifiers in formal and natural languages. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (Vol. 4, pp. 2–132). Dordrecht: Reidel.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science*, *18*, 87–122.