

# The Cognitive Efficacy of Diagrammatic Representations in Logical Reasoning

A Thesis  
Submitted to  
The Graduate School of Letters  
Keio University

In Partial Fulfillment  
of the Requirements for the Degree of  
Doctor of Philosophy

by  
Yuri Sato

February, 2013

# Acknowledgments

I would like to express special thanks to my supervisor, Professor Mitsuhiro Okada, for many years of patient help and fruitful discussions.

I have had the good fortune to receive Professor Atsushi Shimojima's careful support, which was critical for undertaking this interdisciplinary research between logic and cognitive science. Professor Yosaku Nishiwaki and Professor Takashi Iida provided much inspiration from the viewpoint of philosophy.

I am deeply grateful to Dr. Koji Mineshima and Dr. Ryo Takemura for their invaluable advice and constructive discussions.

The experiments were conducted with the help and cooperation of staff at the Keio University GCOE Centre for Advanced Research on Logic and Sensibility, where Professor Shigeru Watanabe and Professor Noburu Notomi provided many opportunities to conduct experiments. Dr. Sayako Masuda and Dr. Yoshiaki Someya generously helped with conducting functional MRI experiments and analyzing the data.

I would also like to acknowledge the kind assistance of Mr. Shigeyuki Nakayama and Mr. Yutaro Sugimoto, who helped me with using  $\text{\LaTeX}$ , and my wife Asumi Sato, who provided several of the illustrations and figures in this thesis.

Finally, I would like to thank my parents Akihiko Sato and Chiharu Sato for their constant encouragement, which enabled me to complete this research.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Diagrammatic reasoning . . . . .	1
1.2	The efficacy of diagrammatic reasoning . . . . .	5
1.3	The questions of effective use of diagrams in logical reasoning . .	10
1.4	The structure of this thesis . . . . .	12
<b>2</b>	<b>Theoretical basis of the efficacy of diagrammatic reasoning</b>	<b>14</b>
2.1	General Hypothesis: Interpretational and Inferential efficacy . . .	14
2.2	Relational analysis of categorical syllogisms . . . . .	19
2.3	Solving syllogistic reasoning with Euler and Venn diagrams . . . .	23
2.4	Predictions . . . . .	37
<b>3</b>	<b>Empirical studies of the inferential efficacy</b>	<b>41</b>
3.1	Reasoning with Euler and Venn diagrams: Experiment 1 . . . . .	41
3.2	Reasoning with Linear diagrams: Experiment 2 . . . . .	53
3.3	fMRI analysis of the efficacy of Euler diagrams: Experiment 3 . .	58
3.4	Children’s reasoning with Euler and Venn diagrams: Experiment 4	71
<b>4</b>	<b>Theoretical–empirical study of the interpretational efficacy</b>	<b>76</b>
4.1	Extracting information from Euler and Venn diagrams . . . . .	76
4.2	Hypothesis: The structural correspondence of interpretations . . .	78
4.3	Sentence-diagram matching test: Experiment 5-1 . . . . .	82
4.4	Diagram-sentence matching test: Experiment 5-2 . . . . .	85
4.5	Discussion . . . . .	89
<b>5</b>	<b>Concluding discussion</b>	<b>90</b>
5.1	Conclusion: The efficacy of diagrams in logical reasoning . . . . .	90
5.2	Further discussion: Structural and procedural correspondence . .	95

5.3	Comparison with related literature . . . . .	98
5.4	Future works . . . . .	103
<b>Appendix A The efficacy of region-based diagrams</b>		<b>105</b>
A.1	Reasoning with Symbolic representations: Experiment 6 . . . . .	105
A.2	Reasoning with Venn diagrams containing three circles: Experiment 7 . . . . .	110
<b>Appendix B Supplemental data of the experiments</b>		<b>120</b>
B.1	Instructions used in the experiments . . . . .	120
B.2	The results of each syllogistic type in the six groups . . . . .	133
<b>References</b>		<b>136</b>

# Chapter 1

## Introduction

This dissertation explores the question about the *cognitive efficacy* of diagrammatic representations in logical reasoning, that is, the question of what makes diagrammatic representations effective for human logical reasoning and what makes the difference between effective and less effective diagrams. Behind this question there is a more general one: How does logical reasoning depend on the way the reasoning is expressed? This dissertation is also intended to make a contribution to a better understanding of the relation between logical reasoning and the way it is represented.

In this chapter, I will outline an approach taken in this dissertation. Section 1 will provide some background on the study of diagrammatic reasoning, discussing various approaches in logic, philosophy, cognitive science, and experimental psychology. I will argue that an interdisciplinary approach is needed to understand the cognitive efficacy of diagrams in logical reasoning. Section 2 will discuss what is distinctive about *logical* reasoning supported by diagrams, in comparison of the role of diagrammatic representations in non-logical problem solving tasks. Section 3 will summarize the main questions discussed in this dissertation.

### 1.1 Diagrammatic reasoning

Humans have a long history of using various types of graphical representation in reasoning and problem solving, such as geometric drawings. Among them there are various kinds of diagrams to support *logical* reasoning, i.e., reasoning with logical connectives and quantifiers. The use of such diagrams has been the subject of continuous research interest over the last several centuries and studied in some depth by logicians and mathematicians. The work of the 17th

century philosopher Leibniz (1903/1988) and subsequent scholars, including Euler (1768), Venn (1881), and Peirce (1897), has been well reported in the literature; in particular, circle diagrams for Aristotelian categorical syllogisms are well known, such as Euler diagrams and Venn diagrams (for the difference between these two diagram types, see Section 3 of Chapter 2).

The dominant view in the field of logic has been, however, that logical reasoning can in principle be expressed in terms of some linguistic or sentential representations; thus, to study logical reasoning and state the validity of logical inference in a precise way, most logicians have been mainly concerned with linguistic or sentential forms of representations rather than with diagrams. This tendency has been particularly manifest since the development of modern logic; although people sometimes rely on diagrams in performing logical and mathematical reasoning, it has been widely held that diagrams themselves are merely auxiliary devices in understanding logical proofs. Thus Tennant (1986) says, in an often quoted passage, “[The diagram] is only an heuristic to prompt certain trains of inference; ... it is dispensable as a proof-theoretic device; indeed, ... it has no proper place in the proof as such (p. 304).” Accordingly, the use of diagrams is often confined to informal introductions to elementary logic including set-theoretical and syllogistic reasoning (e.g., for such classical text books, see Copi, 1953; Allwood, Anderson, & Dahl, 1977; Glymour, 1992).

In contrast to this traditional view, the use of diagrams in logical reasoning has attracted substantial research attention since 1990s. In particular, Barwise and Etchemendy (1991) introduced a formal framework of diagrammatic logic in which the notion of diagrammatic proofs has the same status as linguistic or sentential proofs; in other words, diagrammatically-based reasoning is formalized in the same way that linguistically-based reasoning is formalized in modern logic. We can find a concrete case in Shin’s (1994) seminal work on diagrammatic logic. Shin introduced proof systems for reasoning with Venn diagrams, called Venn-I and Venn-II, where primitive objects constituting a diagram such as circles, points, rectangles and shadings, are considered as formal units and, based on them, semantic and syntactic (inference) rules as corresponding to those in a usual logical system are provided. Since then, various systems of diagrammatic logic have been proposed and studied using the method of mathematical logic (e.g., Howse, Stapleton, & Taylor, 2005). This can be taken as a clear challenge to the critical attitude to diagrams mentioned above. Furthermore, Barwise and

Etchemendy (1986; 1987; 1994) introduced computer-assisted learning systems of logic, using a hybrid interface of logical formulas and diagrams. These studies suggest the possibility of regarding diagrams as objects of formal logical study as well as indispensable units in formal proofs, rather as mere auxiliary, informal devices to support logical reasoning.

In line with the formal study of diagrammatic logic, theoretical researches in the context of cognitive science and artificial intelligence have studied the efficiency or advantage of diagrams in problem solving by comparing between various abstract models of reasoning, in particular, those based on sentential representations and those based on diagrammatic representations. Most notably, Stenning and Oberlander (1995) compared a model of linguistic syllogistic reasoning known as mental model theory in Johnson-Laird (1983) and a model of graphical syllogistic reasoning using Euler diagrams.<sup>1</sup> More relevant to our concern is the seminal study in Shimojima (1996), who presented a theory of the efficacy of representation in general using some semantical notions in logic; in particular, Shimojima provided a detailed analysis of the relationship between a representation and what it represents, discussing the conditions for a representation to be effective in reasoning and problem solving. Comprehensive research into graphical representations came to be referred to as the study of “diagrammatic reasoning” in artificial intelligence and cognitive science (see Narayanan, 1992; Glasgow, Narayanan, & Chandrasekaran, 1995; Allwein & Barwise, 1996 for early collected books).

Currently, however, there seem to be few *experimental* investigations of the cognitive efficacy of diagrams in logical reasoning. Although some experimental studies have examined the effects of diagrams in deductive reasoning, their focus is on elementary reasoning with logical connectives.<sup>2</sup> Hence, little attention has been paid to the cognitive efficacy of diagrams in logical reasoning with sentences involving *quantifiers*, which plays an important role in human deductive reason-

---

<sup>1</sup>The diagrams used in Stenning and Oberlander (1995) are based on a modified version of the traditional system of Euler diagram, and differ from the Euler diagrams adopted in this dissertation. The difference between the two systems will be discussed in Section 3 of Chapter 5.

<sup>2</sup>Thus, Schwartz (1971) and Schwartz and Fattaleh (1972) discussed the case of reasoning with connectives (conjunction, disjunction, conditional, and negation) supported by various types of visual representations (e.g. matrixes and graphs). See Section 2.2 of Appendix 1 for some discussion. And also, Bauer and Johnson-Laird (1993) studied the case of reasoning with “double” disjunction; see Section 1 of Chapter 2 for more details.

ing. In particular, little is known about the cognitive efficacy of Euler and Venn diagrams in syllogistic reasoning, which is a most basic form of quantificational reasoning.

Indeed, Euler diagrams have been the subject of traditional research interest in the context of experimental psychology of syllogistic reasoning. As is well known, performance of syllogistic reasoning has been investigated in various ways since Störing (1908), and a varieties of theories and hypotheses have been proposed to explain how people conduct syllogistic reasoning in experimental settings (for a recent review, see, e.g., Khemlani & Johnson-Laird, 2012). Among them, some theorists, most notably Erickson (1974), proposed that people use a mental analogue of Euler diagrams in solving syllogistic reasoning tasks. More recently, Zielinski, Goodwin, and Halford (2010) discussed the relational complexity analysis of syllogistic reasoning based on the number of regions in Euler diagrams when combining two premise diagrams. In these 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. As exceptional cases, Rizzo and Palmonari (2005) and Calvillo, DeLeeuw, and Revlin (2006) started initial experimental studies on the efficacy of diagrams used as external devices in syllogistic reasoning. However, these studies, as well as most studies in the psychological literature, were based on a traditional system of Euler diagrams, which is known to have some inherent logical problems as we will discuss in Section 3 of Chapter 2. In my view, it is important to take into account the theoretical findings of the recent development of diagrammatic logic and then to test empirically how and whether (a modern version of) Euler diagrams can support syllogistic reasoning with quantified sentences.

As mentioned above, various systems of diagrammatic reasoning, including those for Euler and Venn diagrams, have been proposed in the context of diagrammatic logic (see Stapleton, 2005 for a survey). However, there are few empirical researches to investigate how effective these systems are in people's reasoning. For example, it is often claimed that Venn diagrams are visually less clear and hence harder to handle in actual reasoning than Euler diagrams; but an experimental evaluation of such a claim was seldom provided. Thus, it seems fair to say that a serious gap exists between theoretical and empirical researches in this area.

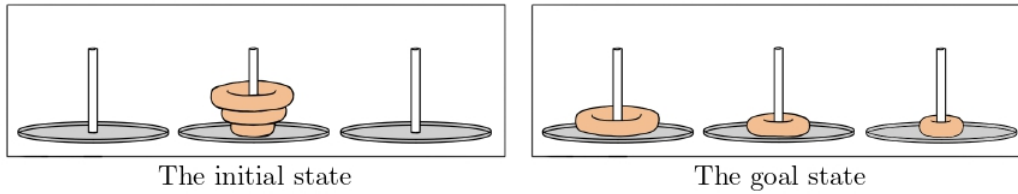
To bridge the gap, this thesis adopts an integrated methodology from the

standpoints of logic, philosophy, cognitive science, and experimental psychology and then discusses the issue on the cognitive efficacy of diagrammatic reasoning, as well as the benefit of using diagrammatic representations in logical reasoning. Before discussing the efficacy of diagrams in logical reasoning, I will consider the role of diagrams in problem solving in general, more specifically, the question of how the ease of solving a problem depends on the way the problem is expressed. This consideration will make clear what is special about logical reasoning and what is common to problem solving in general.

## 1.2 The efficacy of diagrammatic reasoning

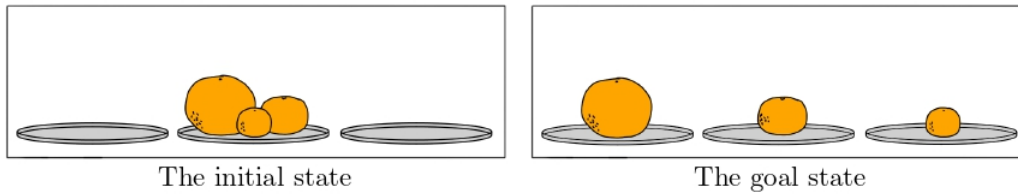
What makes the difference between effective and less effective diagrams for problem-solving? To understanding the role of diagrammatic representations in problem solving, it is important to consider how a representation is matched with a given task. The notion of “match” is emphasized in the empirical framework proposed by Zhang and Norman (1994), which has influenced a number of other researchers in the context of cognitive science and experimental psychology (e.g. Scaife & Rogers, 1996; Norman, 1993). As an illustration, let us take up a typical example of non-logical diagrammatic reasoning with (physical) objects, i.e., the so-called the Tower of Hanoi Puzzle. Zhang and Norman (1994) presented three modified versions of the Tower of Hanoi Puzzle: the Doughnuts Puzzle (Fig. 1.1), the Oranges Puzzle (Fig. 1.2), and the Coffee Cups Puzzle (Fig. 1.3). Physical objects were used in these experiments rather than diagrammatic representations, although this is unlikely to have substantially affected the pattern of results (cf. Scaife & Rogers, 1996). In these tasks, participants are instructed to obey three rules when moving objects (doughnuts, oranges, or coffee cups) from their initial location to a goal location, in order from large–medium–small.

The experimental results have revealed that the Coffee Cups Puzzle is easier than the Doughnuts Puzzle, and the Doughnuts Puzzle is easier than the Oranges Puzzle. This pattern of results can be explained by the number of rules required for successful performance in each task (cf. the checkmarked rules in Figs 1.1, 1.2, and 1.3). In the Oranges Puzzle, all three rules are literally required to solve the puzzle. In contrast, in the Doughnuts Puzzle, participants do not need to be conscious of Rule 3, because it is apparent from the physical properties of the doughnuts. In the Coffee Cups Puzzle, furthermore, because participants



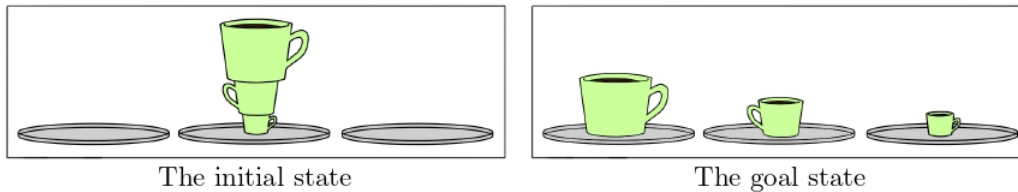
- ✓ Rule 1: Only one doughnut can be transferred at a time.
- ✓ Rule 2: A doughnut can only be transferred to a pole on which it will be the largest.
- Rule 3: Only the largest doughnut on a pole can be transferred to another pole.

Fig. 1.1 The Doughnuts Puzzle (Adapted from Zhang and Norman)



- ✓ Rule 1: Only one orange can be transferred at a time.
- ✓ Rule 2: An orange can only be transferred to a plate on which it will be the largest.
- ✓ Rule 3: Only the largest orange on a plate can be transferred to another plate.

Fig. 1.2 The Oranges Puzzle (Adapted from Zhang and Norman)



- ✓ Rule 1: Only one cup of coffee can be transferred at a time.
- Rule 2: A cup of coffee can only be transferred to a plate on which it will be the largest.
- Rule 3: Only the largest cup of coffee on a plate can be transferred to another plate.

Fig. 1.3 The Coffee Cups Puzzle (Adapted from Zhang and Norman)

are typically careful not to spill the coffee, they tend not to place the smaller cup on the bigger cup, regardless of the instructions. In addition, because of the physical constraints of the task, it is unnecessary for participants to have an explicit understanding of either Rule 2 or Rule 3, to successfully perform the



Coffee Cups Puzzle. Thus, the internal processing of rules in problem solving can be partially *distributed* to the physical constraints of external objects.

The framework of “distributed cognitive tasks” proposed by Zhang and Norman (1994) was used to analyze a relatively simple task, such as the Tower of Hanoi Puzzle, a task that largely depends on *non-declarative* or *procedural* knowledge. Such simple tasks require the knowledge *how* to manipulate the target objects. By contrast, a task of logical reasoning, for example, a task of checking the validity of a logical inference, generally involves processes operating on declarative or propositional knowledge (i.e., *knowing-that*), more specifically, it involves processes of interpreting given sentential premises and obtaining some propositional information, as well as processes of combining the information thus obtained. Accordingly, to discuss the efficacy of diagrams in logical reasoning, one first needs to separate the entire task of logical reasoning into several subtasks and then examine what kind of diagrammatic representation is effective for what kind of tasks involved in logical reasoning. Moreover, in the case of the Tower of Hanoi Puzzle, rules and operations required in solving problems are given to reasoners in a relatively explicit way, as shown above. By contrast, in the case of logical reasoning, rules and operations required to deduce valid conclusions are not entirely explicit; the reasoners must rely on their own implicit knowledge of linguistic expressions and some kind of logical ability. It is not entirely clear how Zhang and Norman’s framework can apply to such complicated cases. Indeed, a negative view of the efficacy of diagrams in inferential processes was 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 more important to deal with reasoning tasks containing multiple processes, and to determine which types of process can be effectively aided by diagrams.

With the distinction between logical and non-logical reasoning tasks in mind, this thesis focuses on categorical syllogistic reasoning with logic diagrams (Euler and Venn diagrams). Although traditional syllogisms are less expressive than standard first-order logic, they are one of the most basic forms of natural language inference, and are thus important in investigating human reasoning (cf. van Benthem, 2008; Moss, 2010).

As an example of (linguistic) categorical syllogisms, consider the possible

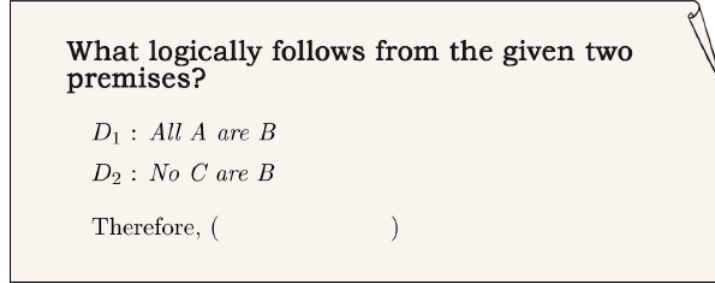


Fig.1.4 An example of a linguistic syllogistic reasoning task

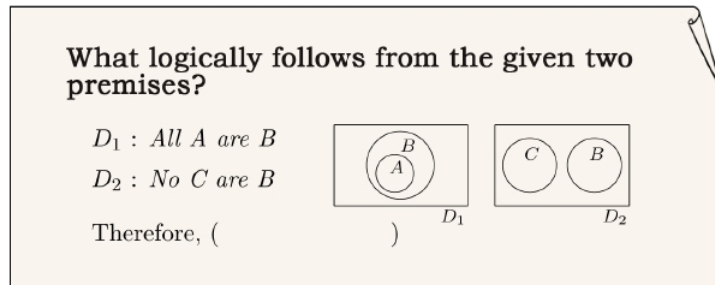


Fig.1.5 An example of a diagrammatic syllogistic reasoning task

process for checking the validity of a syllogism in the task illustrated in Fig. 1.4. Here, the reasoner first interprets the premises *All A are B* and *No C are B*. Then, by integrating the semantic information extracted from quantified sentences, the reasoner judges the logically valid conclusion of the relationship between  $A$  and  $C$ . Note here that, in usual case of logical reasoning tasks, the reasoners are not given instruction regarding strategies (i.e., advice about how to interpret quantified sentences and operate information, or what *logical validity* refers to). As such, the problem solving processes involved in deductive (syllogistic) reasoning tasks are relatively complex. Diagrammatic representations would be expected to be effective even in solving such syllogistic tasks involving complex processes, as observed below. For example, consider the process of checking the validity of a syllogism using Euler diagrams in the task illustrated in Fig. 1.5. The premise *All A are B* is represented by  $D_1$ , and the premise *No C are B* by  $D_2$ . By unifying  $D_1$  with  $D_2$ , as in the right side of Fig. 1.6, one can obtain diagram  $D_3$ . Here the exclusion relation holds between circles  $A$  and  $C$ , from

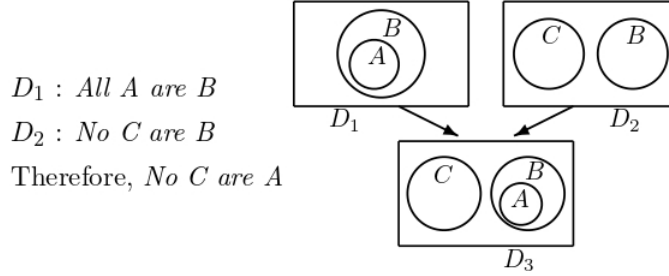


Fig. 1.6 A derivation of syllogistic reasoning with Euler diagrams

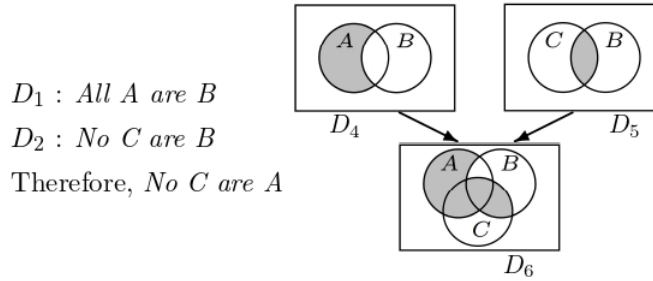


Fig. 1.7 A derivation of syllogistic reasoning with Venn diagrams

which one can extract the correct conclusion “No  $C$  are  $A$ ”. By manipulating diagrams, one can *automatically read off* the correct conclusion. This kind of Euler diagrammatic manipulation was given as a typical example of efficacy in the logical-theoretical framework proposed by Shimojima (1996) mentioned in Section 1. By contrast, Venn diagrams are expected to be more difficult to manipulate syntactically. For example, it would be difficult to combine diagrams  $D_4$  and  $D_5$  of Fig. 1.7 in a direct manner so as to obtain the conclusion diagram  $D_6$ . To make clear the difference in inferential processes between Euler and Venn diagrams, this thesis aims to advance the logical and cognitive study of efficacy presented in Shimojima, by incorporating a proof-theoretical analysis of syllogistic and diagrammatic reasoning based on Mineshima, Takemura, and Okada (2012a,b). Furthermore, I will support my robust claims by connecting the theory of efficacy to empirical findings about syllogistic reasoning with Euler diagrams.

### 1.3 The questions of effective use of diagrams in logical reasoning

The questions discussed in this dissertation can be classified into sub-questions (i)–(vi) below. The theoretical part of this dissertation (Chapter 2) is mainly devoted to the questions (i) and (ii). The questions (iii)–(vi) are discussed in the empirical part of this dissertation (Chapter 3).

(i) As stated in the last section, deductive reasoning tasks involve several processes. The primary question of concern is: for what processes are diagrams effective? There are two possibilities: first, diagrammatic representations may support correct interpretations of sentences. For example, in Fig. 1.6, from the diagram  $D_1$ , one can easily understand that *All A are B* is not logically equivalent to *All B are A*. In addition, diagrammatic representations may improve logical inference. In the case of the syllogism in Fig.1.6, by combining two premise diagrams  $D_1$  and  $D_2$  and observing the topological relationship between the circles in  $D_3$ , one can easily obtain the correct conclusion *No C are A*. Thus, one can differentiate the efficacy components by appealing to the multiple-stage procedures in reasoning. I will deal with this question in Section 1 of Chapter 2.

(ii) Why are certain diagrams effective in interpreting sentential premises and in combining the premise information? This is an explanatory question concerning the conditions required for diagrams to be effective in interpretation and inference as stated in (i) above. Based on semantic and syntactic (proof-theoretical) analyses of syllogistic reasoning with sentences and diagrams I will explain that a *relational structure* is shared as a fundamental structure of syllogistic sentential reasoning and Euler diagrammatic reasoning. The hypothesis to be explored is that Euler diagrams are effective in syllogistic reasoning by virtue of the fact that they make explicit the relational structures implicit in quantified sentences. I will discuss this question in Sections 2 and 3 of Chapter 2. Furthermore, to clarify conditions for diagrams to be effective in interpretational processes, theoretical and empirical aspects of information extraction processes from logic diagrams are discussed in Chapter 4.

(iii) What types of diagram are effective in actual processes of solving syllogistic reasoning tasks? As I mentioned in the previous section, the two types of diagrams, namely Euler diagrammatic type and Venn diagrammatic type, are considered to be important in the literature. Hence the above question can be

stated: which out of the two types (i.e., Euler diagrammatic type and Venn diagrammatic type) are more effective? To address the above question in a way that reflects people’s actual performance, through an experiment, I will present an experimental study of which types of logic diagrams actually aid untrained people in conducting deductive reasoning. This is the main empirical question that is directly relevant to the theoretical questions in (i) and (ii). I will deal with this question in Section 1 of Chapter 3.

(iv) If the claim in (ii) above is correct, it would be expected that any diagram that makes explicit the relational information of a categorical sentence in a suitable way would be effective in supporting syllogistic reasoning. For example, if a diagram composed of two-dimensional objects (such as circles) is effective, does the same hold true for one-dimensional versions of diagrams (i.e., what we call *linear* diagrams)? The question of whether linear diagrams can be effective in syllogistic reasoning has seldom been investigated experimentally. The study of linear diagrams supplements the experiment in (iii) and hence provides further empirical support to the theoretical claim made in (ii). The question (iv) will be dealt with in Section 2 of Chapter 3.

(v) What is the neural substrate underlying diagrammatic reasoning? I will discuss which brain areas are activated during the use of diagrams in reasoning tasks, using functional magnetic resonance imaging (fMRI) to examine brain-hemodynamic changes. Many neuroimaging studies have focused on deductive sentential reasoning (for a review, see Goel, 2007). However, there have been few attempts to apply these techniques to the study of deductive diagrammatic reasoning. A study using a neuroimaging technique will provide further support to the theoretical model of diagrammatic reasoning presented in (i), in particular, the distinction between interpretational and inferential efficacy. I will deal with this question in Section 3 of Chapter 3.

(vi) Is the manipulation of Euler-diagrams intuitive and easy-to-understand? To approach this question, it is interesting to see how beneficial diagrams are in children logical reasoning, because children seem to have small amount of prior knowledge concerning Euler and Venn diagrams. The cognitive efficacy of Euler diagrams in children’s syllogistic reasoning will be examined in Section 4 of Chapter 3 (see also Section 1 of Chapter 5 for a relevant discussion).

## 1.4 The structure of this thesis

The structure of this thesis is as follows. First, in Chapter 2, I provide a theoretical framework of the efficacy of logic diagrams in deductive reasoning and present the main hypothesis I will defend in this dissertation. In Section 1, a cognitive model of logical reasoning with sentences and diagrams is introduced to clarify the roles of diagrams in interpretational and inferential processes. I will distinguish two kinds of efficacy, i.e., interpretational and inferential efficacy, and then discuss how to test whether diagrams have interpretational or inferential efficacy. Section 2 presents a logical analysis of categorical syllogisms based on the idea that categorical sentences are analyzed in terms of relational structures (Mineshima, Okada, & Takemura, 2012b). Section 3 presents a relational analysis of syllogistic reasoning with Euler diagrams (Mineshima, Okada, & Takemura 2012a). I will then put forward the hypothesis that the efficacy of Euler diagrams in syllogistic reasoning crucially depends on the fact that they are effective ways to represent and reason about relational structures implicit in categorical sentences. In Section 4, the above consideration is summarized as a form of predictions for the experiments conducted in this dissertation.

Chapter 3 presents an empirical study of efficacy of logic diagrams. Section 1 describes an experiment comparing participants' performance in syllogism solving using Euler diagrams and Venn diagrams. Section 2 examines whether the pattern of results in Experiment 1 also emerges when a linear variant of Euler diagrams is used. Section 3 describes an fMRI experiment comparing hemodynamic changes when Euler diagrams are used in reasoning with cases in which diagrams are not used. In Section 4, children's performance is compared in syllogistic reasoning tasks externally supported by Euler and Venn diagrams. The results of the four experiments provide evidence supporting the hypothesis that Euler diagrams have inferential efficacy, that is, they are effective in inferential processes of combining the premise information.

Chapter 4 is concerned with a theoretical and empirical analysis of interpretations of sentences and diagrams. The conditions in which diagrams can have interpretational efficacy are further explored. More specifically, the process of information extraction from Euler and Venn diagrams is investigated via experiments examining the matching time between categorical sentences and the corresponding diagrams.

Chapter 5 discusses the conclusions that can be drawn from the current findings. In Section 1, the findings of each of the present experiments are summarized, and the efficacy of diagrams is assessed by considering the way in which spatial simulations make relational information available. Section 2 considers the correspondence relation between representation systems (what is called *structural* correspondence) and the correspondence relation between inference systems (what is called *procedural* correspondence) as conditions for realizing the two types of efficacy. Section 3 compares the theoretical and experimental findings in this thesis with related studies in cognitive science. Specifically, the graphical method of case identification and the dual process theory of reasoning are discussed. Finally, possible future research directions are suggested and discussed.

In Appendix A, the efficacy of Venn diagrams is discussed. Venn diagrams are characterized as “region-based” diagrams, in sharp contrast to Euler diagrams, i.e., diagrams that are defined in terms of relations. Two experiments are described to examine the interpretational and inferential effects of Venn diagrams. Section 1 presents an experiment comparing between syllogistic reasoning with Venn diagrams and reasoning with set-theoretical symbolic representations corresponding to Venn diagrams. Section 2 presents an experiment using Venn diagrams with three circles (3-Venn diagrams) instead of those with two circles (2-Venn diagrams). The results of the two experiments provide support for the claim that Venn diagrams have interpretational efficacy, and that a special version of Venn diagrams, 3-Venn diagram, has inferential efficacy. Furthermore, I discuss the possibility of a reasoning task (other than traditional categorical syllogisms) which is more appropriately matched to Venn diagrams.

Appendix B provides supplemental data for the experiments conducted in this thesis.

## Chapter 2

# Theoretical basis of the efficacy of diagrammatic reasoning

### 2.1 General Hypothesis: Interpretational and Inferential efficacy

In this section, I consider how external diagrams can affect deductive reasoning. In accord with cognitive modeling, there are two ways in which diagrams can be effective in deductive reasoning. As such, a methodology can be introduced to empirically distinguish the two kinds of efficacy.

Typical examples of deductive reasoning problems using external logic diagrams are shown in Fig. 2.1.

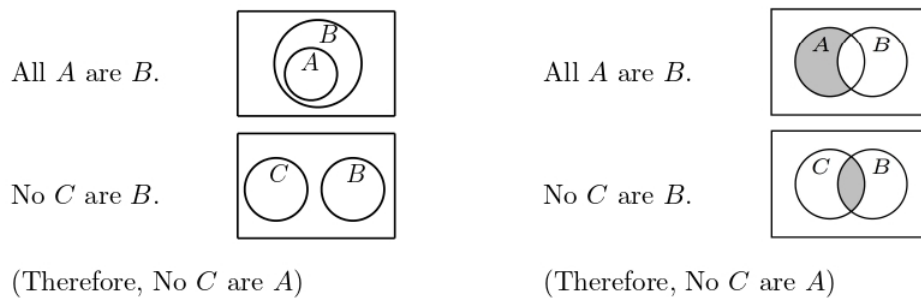


Fig. 2.1 Examples of syllogistic reasoning tasks with diagrams

Here a syllogism is presented with two types of logic diagram (Euler and Venn diagrams). How can such diagrams contribute to checking the validity of a deductive argument? First, consider a cognitive model of deductive problem solving



with diagrams, as shown in Fig. 2.2. This model highlights two possible roles of diagrams in deductive reasoning.

Sentential reasoning assumes a standard two-stage framework in natural language semantics (see, e.g. Blackburn & Bos, 2005), according to which sentences are first associated with semantic information. The validity of the argument is then checked using inferential mechanisms (such as model-theoretical or proof-theoretical mechanisms). The details and precise nature of such linguistic comprehension and inference are outside the scope of this study.

Diagrams are also associated with semantic information, but in this case constitute syntactic objects to be manipulated in reasoning processes. Two ways in which diagrams can be effective in deductive reasoning can be distinguished. It should be noted that this model does not exclude other possible factors contributing to efficacy, including the self-construction process of diagrams (see Chapter 3, Section 1.3).

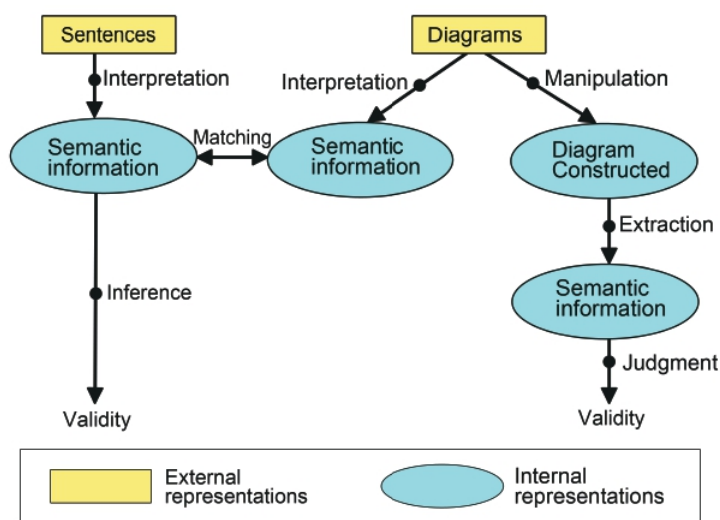


Fig. 2.2 Cognitive model of deductive reasoning with diagrams

**Interpretational efficacy.** First, diagrams can help to improve interpretations of sentences, thereby avoiding deductive reasoning errors due to misinterpretation. 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 & Chapman, 1959). However, participants presented

with diagrams like those in Fig. 2.1 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)symmetry of diagrams in Stenning, 2002). In the current model, 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 & Etchemendy, 1991).

**Inferential efficacy.** Second, diagrams can play a crucial role in inference processes themselves. For example, the process of solving deductive reasoning tasks can be replaced by manipulation of diagrams, constructing “diagrammatic reasoning” to check the validity of a deductive argument. The above model assumes 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.

Such manipulations are typically considered to be triggered by external diagrams and carried out internally, without actually drawing or moving physical objects. However, the use of such internal manipulations of diagrams has been the subject of controversy (see, e.g. Schwartz, 1995; Scaife & Rogers, 1996). Indeed, it is widely held that diagrams can serve as memory aids or auxiliary sources of information in deductive problem solving. As stated above, Larkin and Simon (1987) argue that *inference* (which is distinguished from *search* and *recognition*) is largely independent of the way in which information is represented, implying that diagrams provide minimal benefit in terms of inference when mechanical (pulley) and geometric problem solving processes are formalized as computational models. Bauer and Johnson-Laird (1993) examined the efficacy of circuit and jigsaw diagrams in deductive reasoning with double disjunction, arguing that the diagrams are often used to keep track of alternative models, as postulated in Johnson-Laird’s mental model theory. Behind their approach there is an assumption that diagrams are visualizations of *models*, rather than *syntactic representations* indicating models. Thus, it has been a traditional view that reasoning depends on the semantics but little depends on the way the reasoning is expressed.

Given the conflicting findings discussed above, it is important to investigate whether internal manipulations of diagrams are used during actual reasoning with

diagrams. Hegarty (1992) reported evidence for the existence of “mental animation”, based on reaction times and eye fixations during ensuing events as a result of motion in kinematic problem solving involving simple mechanical systems (the pulley problem). Trafton, Trickett, and Mintz (2005) observed the thought processes of scientists (astronomers and physicists) while conducting their own research using scientific visualizations, by coding the number of “spatial transformations”, or the mental alignment of external-internal diagrams or visualizations. Shimojima and Fukaya (2003) and Shimojima and Katagiri (2008a,b) argue for the existence of “inference by hypothetical drawing” involving imaginary drawing of the given diagrams, based on eye-tracking data of participants working with position diagrams in transitive inference tasks.

In this thesis, I will focus on complex deductive reasoning tasks, namely, syllogistic reasoning tasks, and on the effects of logic diagrams externally given therein. I will present evidence for the existence of diagram manipulation on the basis of experiments comparing participants’ performance in syllogism solving tasks where logic diagrams of several different forms are given. My claim is consistent with the influential view in the study of external representations in general, namely, that (a) external representations can be used without being interpreted, and that (b) they can change the nature of tasks, namely, tasks with and without external representations are completely different from users’ point of view (see Zhang & Norman, 1994, mentioned above).

The general hypothesis I will defend in this thesis can be stated as follows:

**Hypothesis on the inferential efficacy of diagrams.** Diagrams can be effective in supporting deductive reasoning if the solving processes (i.e. processes of drawing a conclusion) are replaced by processes composing the manipulation of diagrams.

More specifically, it is claimed that (1) logic diagrams can have inferential efficacy, that is, syntactic manipulation of diagrams occurs in deductive reasoning with external logic diagrams, and that (2) certain diagrams would naturally trigger the constructions of diagrammatic reasoning so that even users without explicit prior knowledge of inference rules or strategies could correctly manipulate such diagrams.

**Methodology.** One way to test the above claims is to compare performance in deductive problem solving with several distinct diagrams that are equivalent in

semantic information differ in form, specifically comparing diagrams with forms that are suitable for syntactic manipulation and diagrams with forms that are not. To this end, the current research focused on two types of logic diagrams, Euler and Venn diagrams. This research is supported by the basic assumption that the existence of syntactic manipulation of diagrams depends on the forms of the diagrams used, and the simplicity or naturalness of the required diagrammatic reasoning. Thus, if performance is significantly improved by the use of diagrams of a form suitable for diagrammatic reasoning constructions, then this finding would support the existence of such constructions in human reasoning.

To test the claims described in (2), the participants in my experiments were presented with instructions describing the meaning of the categorical sentences and diagrams used, but were not given any instruction about the rules or strategies for the syntactic manipulation of the diagrams. I expect that if certain diagrams have inferential efficacy, this will be exploitable based on their natural properties or constraints, rather than extra conventions. In other words, processes of syntactic manipulation of diagrams may be conducted without an explicit knowledge of the underlying rules or strategies, as postulated in my cognitive model. Indeed, diagrammatic reasoning constructions may be naturally triggered based on a correct understanding of the meaning of specific diagrams.

The efficacy of logic diagrams has been investigated in the context of studies of logic teaching method (Stenning, Cox, & Oberlander, 1995; Monaghan & Stenning, 1998; Dobson, 1999; Eysink, Dijkstra, & Kuper, 2001). In these studies, participants are typically provided with substantial training about different ways of manipulating diagrams. In contrast, in the current studies examined the question of whether diagrams can be useful for people who are not trained in rules or strategies of diagrammatic deductive reasoning. This question is important because, in contrast to logical formulas in symbolic logic, logic diagrams in general are considered more intuitive and effective for reasoning among novices, rather than for expert or machine reasoning. In view of the complexities of the problem solving processes of deductive reasoning (e.g., Levesque, 1988), it should be asked whether logic diagrams exhibit this property. Given that diagrams are commonly used in logic teaching, they should be easily learnable, even by people who are untrained in the conventions governing the representational systems of the given diagrams. If mastery of diagrammatic representations is only possible through substantial training, then it would clearly require some knowledge about

certain basic notions of logic, and hence, one would need to learn logic before the use of diagrams. In Plato’s *Meno*, Meno’s slave answered a series of Socratic questions and came to understand the general theorem of geometry via the use of diagrams (cf. Plato’s *Meno*; the relevant passage is examined in Giaquinto, 1993), suggesting that the question of which logic diagrams are learnable without substantial training is important in the context of logic teaching.

In light of the above general hypothesis, I will investigate the efficacy of Euler and Venn diagrams in supporting *sylogistic* reasoning. To this end, building on Mineshima, Okada, and Takemura (2012a,b), I will first present a framework in which both syllogistic and diagrammatic inferences are analyzed in a unified way, referred to as a “relational perspective”. The key assumption is that both types of inferences are decomposed as inferences with two primitive relations (i.e., inclusion and exclusion). I propose that the efficacy of Euler diagrams in syllogistic reasoning arises from their effectiveness in representing and reasoning about relational structures that are implicit in quantified sentences. In Section 2, I present on a relational analysis of categorical sentences and syllogisms. In Section 3, I turn to the relational analysis of Euler diagrams. Finally, in Section 4, these claims are summarized as predictions for empirical experiments.

## 2.2 Relational analysis of categorical syllogisms

In this section, I will analyze the interpretations of categorical sentences and inferences of categorical syllogisms, based on the work of Mineshima et al. (2012b). I will then propose an ideal (based on maximum competence) process of interpretation and inference with respect to categorical syllogisms. Given the analysis, I will then provide a hypothesis on the efficacy of Euler diagrams in syllogistic reasoning.

### 2.2.1 Categorical syllogisms decomposed

This section provides an overview of an abstract formulation of syllogistic inferences using relational notations, as far as it is relevant to experimental discussion. The two types of relation, inclusion  $\sqsubset$  and exclusion  $\sqsupset$ , are relatively simple, but are expressive enough to represent a categorical syllogism. Mineshima et al. (2012b) refer to an abstract symbolic inference system based on these two kinds of relations a *generalized syllogistic inference system*, abbreviated as GS.

The categorical sentences used in a syllogism contain quantifiers such as *all*, *some*, *some-not*, and *no*. Within standard view of modern logic, such sentences are formalized in first-order logic, which deals with quantification over individuals. For example, *All A are B* in natural language is translated into  $\forall x(Ax \rightarrow Bx)$ , namely, “for every  $x$ , if  $x$  is  $A$  then  $x$  is  $B$ ”. In the context of natural language semantics (e.g., Montague, 1974; Barwise & Cooper, 1981), in contrast, such quantified sentences are treated by stating relations between sets. Interestingly, this set-theoretical view of quantified sentences was also proposed as a real human interpretation (i.e., mental representations) in the literature of cognitive psychology (e.g., Johnson-Laird, 1983; Geurts, 2003a; Politzer, van den Henst, Luche, & Noveck, 2006). Building on this view, in the **GS** system introduced in Mineshima et al. (2012b), *All A are B* can be analyzed as having a logical form  $A \sqsubset B$  ( $A$  is included in  $B$ ), which semantically expresses that  $\mathbf{A} \subseteq \mathbf{B}$ . Similarly, *No A are B* is analyzed as having a logical form  $A \sqsupset B$  ( $A$  is excluded in  $B$ ), expressing that  $\mathbf{A} \cap \mathbf{B} = \emptyset$ . In this way, the semantic primitives of quantificational sentences can be regarded as the relations between sets, such as subsets and disjointness. The translation of a categorical sentence into a **GS**-formula is summarized as the following “Interpretation Rules”:

**Interpretation Rules:**

1. *All A are B* is decomposed into “ $A$  is included in  $B$ ” ( $A \sqsubset B$ ).
2. *No A are B* is decomposed into “ $A$  is excluded in  $B$ ” ( $A \sqsupset B$ ).
3. *Some A are B* is decomposed into “ $c$  is included in  $A$  and  $c$  is included in  $B$ ” ( $\{c \sqsubset A, c \sqsubset B\}$  for some  $c$ ).
4. *Some A are not B* is decomposed into “ $d$  is included in  $A$  and  $d$  is excluded in  $B$ ” ( $\{d \sqsubset A, d \sqsupset B\}$  for some  $d$ )."

The crucial point here is that existential sentences are *decomposed* or *exposed* in terms of inclusion and exclusion (see Interpretation Rules 3 and 4). This translation is similar to Aristotle’s formulation in *ecthesis*, with a philosophical background that is described in Smith (1982), and psychological reality that was studied by Politzer and Mercier (2008) and Politzer (2011); see references to modern logic studies on *ecthesis* given there.

Given this translation, all the valid inferences in categorical syllogism can be transformed into ideal solving processes based on the **GS** proof system in

Mineshima et al. (2012b). The crucial inference rules in the inference system are involving the inclusion and exclusion relations, labeled as ( $\sqsubset$ ) and ( $\vdash$ ):

**Inference Rules:**

**Inclusion ( $\sqsubset$ ) Rule**

If  $X$  is included in  $Y$  and  $Y$  is included in  $Z$ , then  $X$  is included in  $Z$ .

**Exclusion ( $\vdash$ ) Rule**

If  $X$  is included in  $Y$  and  $Y$  is excluded in  $Z$ , then  $X$  is excluded in  $Z$ .

Consider the following examples of how to check the validity of categorical syllogisms using the inclusion and exclusion rules. To begin with, a syllogism *All A are B, No B are C; therefore No A are C* can be represented as a simple ideal solving process as shown in Fig. 2.3. The inference rule used here is the only Exclusion ( $\vdash$ ) Rule.

1. *All A are B.* (Premise 1)
2. *No B are C.* (Premise 2)
3.  $A$  is included in  $B$  ( $A \sqsubset B$ ). (By Interpretation Rule 1 from 1)
4.  $B$  is excluded in  $C$  ( $B \vdash C$ ). (By Interpretation Rule 2 from 2)
5. Hence,  $A$  is excluded in  $C$  ( $A \vdash C$ ). (By Exclusion Rule from 3 and 4)
6. Therefore, *No A are C.* (By Interpretation Rule 2 from 5)

Fig. 2.3 Derivation of *No A are C* from *All A are B* and *No B are C*

As an example involving an existential sentence, consider an ideal solving process for the syllogism *Some A are B, All B are C; therefore Some A are C*, as shown in Fig. 2.4. It should be noted here that steps 4 and 7 use the inference rules corresponding to **elimination of conjunction** in standard natural deduction system; step 8 uses the rule corresponding to **introduction of conjunction**.

All valid categorical syllogisms (with and without existential import) can be simulated in **GS**. This means that syllogistic inferences can be decomposed as a kind of relational inferences about inclusion and exclusion, i.e., a higher order relational inference concerning sets rather than individuals.

1. *Some A are B.* (Premise 1)
2. *All B are C.* (Premise 2)
3. *d* is included in *A* ( $d \sqsubset A$ ) and *d* is included in *B* ( $d \sqsubset B$ ). (By Interpretation Rule 3 from 1)
4. Hence, *d* is included in *B* ( $d \sqsubset B$ ). (From 3)
5. *B* is included in *C* ( $d \sqsubset C$ ). (By Interpretation Rule 1 from 2)
6. Hence, *d* is included in *C* ( $d \sqsubset C$ ). (By Inclusion Rule from 4 and 5)
7. *d* is included in *A* ( $d \sqsubset A$ ). (From 3)
8. Hence, *d* is included in *A* ( $d \sqsubset A$ ) and *d* is included in *C* ( $d \sqsubset C$ ). (From 6 and 7)
9. Therefore, *Some A are C.* (By Interpretation Rule 3 from 8)

Fig. 2.4 Derivation of *Some A are C* from *Some A are B* and *All B are C*

### 2.2.2 Hypothesis on inferential efficacy of Euler diagrams

If the relational information encoded by categorical sentences was transparent to untrained reasoners, it would be much easier for them to solve categorical syllogisms. However, cognitive psychological studies of deductive reasoning accumulated suggest that this is not the case. For example, logically untrained people often interpret *All A are B* as equivalent to *All B are A*, indicating that the relational information  $A \sqsubset B$  is not directly available to them. Similarly, the difficulties observed in solving categorical syllogisms involving existential sentences (cf. chap. 7 of Evans, Newstead, & Byrne 1993 for a survey of relevant experimental data) suggest that there is a certain gap between ordinary ways of performing existential inferences and relationally decomposed processes, as described above.

Taking the real semantics of sentences into account raises the problems of pragmatic interpretations. In the case of sentences appearing in syllogisms, potential factors affecting the interpretations include word order (e.g. “illicit conversion errors” mentioned before), information structure (e.g. topic-focus articulation), and varieties of pragmatic inferences (in particular, scalar implicatures triggered by quantifiers; e.g., *Some A are B* tends to implicate *Not all A are B* and hence *Some A are not B*). See Stenning (2002) for an overview of these issues. Relational inferences are important to keep track of the *direction* of a relation, for



example, whether  $A \sqsubset B$  or  $B \sqsubset A$  holds. However, it has been well established that the direction of a relation that has been encoded by a quantified sentence can sometimes be obscured by non-semantic factors (see Oberauer, Hönig, Weidenfeld, & Wilhelm, 2005; Oberauer & Wilhelm, 2000). As such, studying the processes of inference requires a method for fixing the interpretations of natural language sentences, then extracting the purely semantic information. In these respects, Euler diagrams can be effective tools; Euler diagrams can be used to visualize the Interpretation Rules of categorical sentences and Euler diagrammatic manipulation can be used to visualize the Inference Rules such as Inclusion and Exclusion Rules.

As mentioned in Section 1, there are two ways in which diagrams can externally support ordinary reasoning. In the case of syllogistic reasoning, I will hypothesize the following:

**Hypothesis on the efficacy of Euler diagrams in syllogistic reasoning.** Euler diagrams can be effective in supporting syllogistic reasoning because they can make explicit the relational information (i.e., inclusion and exclusion relations) contained in categorical sentences, thereby replacing the process of drawing a valid conclusion with the process of manipulating diagrams.

In the next section, I will turn to an analysis of Euler diagrammatic reasoning.

## 2.3 Solving syllogistic reasoning with Euler and Venn diagrams

In this section, I will first describe several diagrammatic representations used in logical reasoning tasks. In particular, I will focus on the modified Euler diagrams (EUL), based on the works of Mineshima, Okada, Sato, and Takemura (2008) and Mineshima, Okada, and Takemura (2012a). Second, I describe potential problem solving processes using categorical syllogisms with diagrams. The two types of possible efficacy (interpretational and inferential efficacy) are discussed in Section 3.1 and 3.2, respectively.

### 2.3.1 Diagrammatic representation systems used to depict partial information

As emphasized in Stenning and Oberlander (1995), diagrams that are beneficial as a tool in deductive 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 reasoning 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. Some previous research, particularly studies of the psychology of reasoning, has not distinguished between semantic and pragmatic layers in interpretations of premise sentences. On the other hand, logic studies have developed methods of separating the two layers and describing semantic layers. Thus, logical studies of diagrammatic representations are worthy of research attention, enabling the observation and performance of inference with literal controlled interpretations, or inference with purely semantics.

An Euler diagrammatic representation system introduced in our previous work (Mineshima et al., 2008; Mineshima et al., 2012a, where a formal semantics and a diagrammatic inference system are provided for it), called the **EUL** system, can serve as such a system. Hence, of the various representation systems based on Euler diagrams, I use the **EUL** system in the experiments reported in this thesis. Here let us briefly explain the **EUL** system in comparison with another type of Euler diagrammatic representation systems (Gergonne’s system) and Venn representation system.

**Problems with traditional logic diagrams.** Leonhard Euler (1768) introduced his diagrams to teach Aristotelian syllogistic logic to a German princess.<sup>1</sup>

---

<sup>1</sup>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 13th century scholar Ramon Lull (1617). Furthermore, in Leibniz’s (1903/1988) work in the 17th century, there is the description of the use of diagrams to represent syllogisms although it was only much later that his work was published.

As mentioned in Chapter 1, Euler diagrams represent set relationships in terms of the spatial relationship (inclusion and exclusion relations) between circles. For example, universal sentences of *All A are B* is represented in terms of the inclusion relation between circles as in  $D_1^e$  on the left side of Fig. 2.5. Similarly, *No A are B* is represented in terms of the exclusion relation between circles as in  $D_2^e$  on the right side of Fig. 2.5.

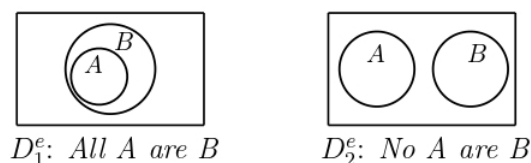


Fig. 2.5 Correspondence between universal sentences and Euler diagrams.

With respect to existential sentences, several types of diagram have been proposed in Euler's (1768) work (see also Hammer & Shin, 1998). Among them, there is a particular version of Euler representation systems, which we call Gergonne's representation system (Gergonne, 1817; Kneale & Kneale, 1962, 349–352), a system which is well known in the context of the psychological studies of syllogistic reasoning (e.g., Erickson, 1974). The features of Gergonne's system are stated as follows: (i) diagrams consist only of circles; (ii) every minimal region (namely, a region having no other region contained within it) in a diagram represents a non-empty set. As a consequence, diagrams in Gergonne's system cannot represent partial (i.e., disjunctive) information in a single diagram. For example, to represent an existential sentence in syllogisms, one has to use more than one diagram. Thus, the sentence *Some A are B* is represented by the *disjunction* of four diagrams, namely, diagrams of the following forms (see Fig. 2.6): (1) circle *A* is inside circle *B*, (2) circle *B* is inside circle *A*, (3) circle *A* and circle *B* coincide, and (4) circle *A* partially overlaps with circle *B*. Similarly, *Some B are not C* requires three diagrams.

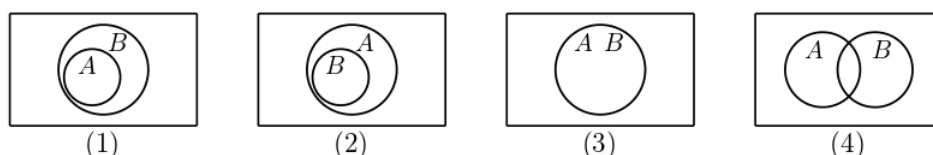


Fig. 2.6 Traditional Gergonne's diagrams corresponding to *Some A are B*.

This situation of multiple diagrams means that to check the validity of a syllogism with these two premises, one has to take into account twelve ways of combining diagrams (cf. chap. 4 of Johnson-Laird, 1983, where this was mentioned as a problem on “combinatorial explosion” of premise diagrams). Thus, although the diagrams in Gergonne’s system have visual clarity in that they solely rest on the topological relationships between circles, they cannot represent partial information (which is essential to the treatment of syllogistic reasoning) in a single diagram, and hence, they are difficult to handle in actual deductive reasoning. Thus, Gergonne’s system is relatively intuitive, but lacks a structured mechanism to deal with abstraction. The Euler diagrams used in the experiments of Rizzo and Palmonari (2005), Calvillo, DeLeeuw, and Revlin (2006), and Roberts and Sykes (2005) mentioned in Chapter 1 can also be considered to be based on this version of the Euler diagrammatic representation system (namely, Gergonne’s diagrams). 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, I will focus on the diagrams avoiding the defectiveness. In fact, the Euler diagrams currently studied in diagrammatic logic are typically based on the convention of crossing, rather than the type used in Gergonne’s system (cf. Stapleton, 2005).

**Venn diagrams.** Venn (1881) and Peirce (1897) attempted to overcome the shortcomings of Gergonne’s version of the Euler diagram by removing the existential import from regions (for historical background see Baron, 1969; Gardner, 1958; Greaves, 2002). Venn (1881) focused on “primary diagrams”, such as the one in Fig. 2.7, where circles partially overlap each other.

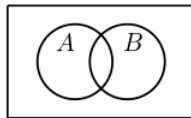


Fig. 2.7 Partially overlapping circles

Partially overlapping circles as in Fig. 2.7 are semantically vacuous (i.e., they deliver no semantic information about the relation between the sets denoted by the circles). Accordingly, they can be used to express that the semantic relationship between the circles is *indeterminate*. Thus, Venn diagrams are based on the convention of *crossing*, according to which two circles which are indeterminate

with respect to their semantic relation partially overlap each other. As a symbolic notation, we use  $\bowtie$  to indicate a *crossing* relation between two circles as in Fig. 2.7. Thus, in Venn diagrams, meaningful relations among circles are expressed using novel syntactic devices.

First, 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, then represented as in diagrams  $D_1^v$  and  $D_2^v$  of Fig. 2.8. In this way,

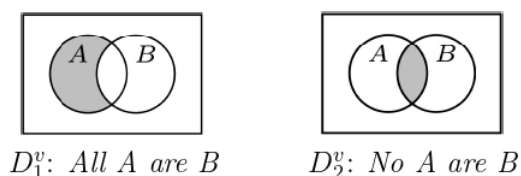


Fig. 2.8 Correspondence between universal sentences and Venn diagrams.

logical relations among terms are represented not simply by topological relations between circles, but by the essential use of shading.

Second, 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. 2.9.

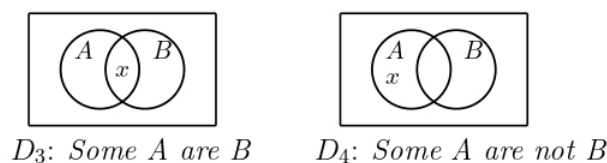


Fig. 2.9 Correspondence between existential sentences and Euler diagrams.

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. Thus, the expressiveness of Venn diagrams is obtained at the cost of intuitiveness of diagrammatic representations.

**Euler (EUL) diagrams.** The EUL system is a simple representation system combining the features of the original version of Euler diagrams (Gergonne’s system) and Venn diagrams.<sup>2</sup> Following traditional Euler graphical systems, 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. To avoid the complexity of Gergonne’s system, the EUL system adopts the convention of crossing and uses a named *point* “ $x$ ” to indicate the existence of objects. Syllogistic reasoning can be characterized in terms of an inference system based on EUL (cf. Mineshima et al., 2008; Mineshima et al., 2012a). 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 convention of crossing. In what follows, we refer to diagrams in the EUL system simply as Euler diagrams.

In the EUL representation system, there are three types of relations that can be distinguished: (i) a circle or a point  $x$  is located inside a circle  $A$ , symbolically written as  $x \sqsubset A$ ; (ii) a circle or a point  $x$  is located outside a circle  $A$ , written as  $x \sqsupset A$ ; and (iii) a circle  $A$  and a circle  $B$  partially overlap each other, written as  $A \bowtie B$  (see Fig. 2.7). In this symbolic notation, we use the same binary symbols,  $\sqsubset$  and  $\sqsupset$ , as in GS (see also Section 2), for the relations in (i) and (ii). Here it is easily seen that such abstract representation of diagrammatic relations can be translated into the formulas in GS. As described in more detail below, given the relational formalization of Euler diagrams, both linguistic and diagrammatic syllogistic inferences can be represented as a type of *relational* inference concerning sets. This enables the formulation of relevant inference rules underlying syllogistic inference in a relatively simple way: all the relevant inference rules are those concerning the transitivity of  $\sqsubset$  and  $\sqsupset$ . For more details on this point, see Section 3.2.

Now the relationships between categorical sentences, Euler diagrams, and

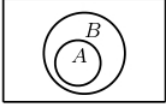
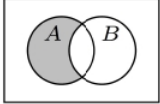
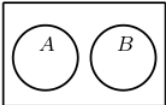
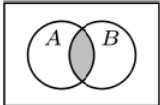
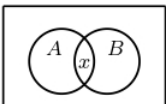
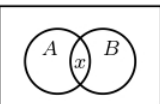
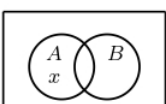
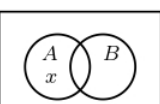
---

<sup>2</sup>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.

Venn diagrams, as well as abstract representations of them are summarized in Table 2.1. It is important to note that the *structural correspondence* holds between the interpretations of Euler diagrams and categorical sentences. To put the assertion more concretely, in the first line of Table 2.1, the categorical sentence *All A are B* is naturally interpreted as a relational representation such as  $\mathbf{A} \subseteq \mathbf{B}$ , which denotes the subset relation between sets  $A$  and  $B$ . The relevant Euler diagram can also be interpreted as a relational representation, such as  $\mathbf{A} \subseteq \mathbf{B}$ . In this case, the *structural correspondence* between these representations clearly holds. When the Euler diagrams are used to interpret categorical sentences, it is expected that some pragmatic interpretations of categorical sentences will be avoided, meaning that the pure semantics of the sentences (i.e., Interpretation Rules) can be examined.

On the other hand, in the relevant Venn diagram, the set relation is expressed using shading, with shaded regions denoting empty sets. Thus, the Venn diagram is directly interpreted as follows; the intersection between  $A$  and the complement

Table 2.1 Representations of categorical sentences in Euler and Venn diagrams.

Categorical sentences	Euler diagrams (Symbolization in EUL)	Venn diagrams	Semantics
All $A$ are $B$	 $A \sqsubset B$		$\mathbf{A} \subseteq \mathbf{B}$
No $A$ are $B$	 $A \sqcup B$		$\mathbf{A} \cap \mathbf{B} = \emptyset$
Some $A$ are $B$	 $x \sqsubset A, x \sqsubset B$		$\mathbf{A} \cap \mathbf{B} \neq \emptyset$
Some $A$ are not $B$	 $x \sqsubset A, x \sqcup B$		$\mathbf{A} \cap \overline{\mathbf{B}} \neq \emptyset$

of  $B$  is the empty set  $\{\mathbf{A} \cap \overline{\mathbf{B}} = \emptyset\}$  “there is nothing which is  $A$  but not  $B$ ”. Since this is a paraphrased form of the above interpretation of a categorical sentence, the two representation systems are semantically equivalent. However, in comparison to the Venn diagrams, the *structural correspondence* between representational elements does not hold in the current sense. Based on this view, the issue of correspondence between representational elements is discussed in more detail in Chapter 4. Let us next attempt to extend the observation to the analysis of the inferential efficacy. As stated in Section 1, some comparison between reasoning using Euler diagrams and Venn diagrams would be intended to distinguish inferences from interpretations.

### 2.3.2 Checking the validity and invalidity of categorical syllogisms using diagrams

Given the correspondence between Euler diagrams and categorical sentences, the process of solving categorical syllogisms can be replaced by the process of manipulating Euler diagrams, particularly unifying premise diagrams, and extracting information from them. As shown in Section 1, logic diagrams can help a reasoner to solve syllogistic reasoning tasks at both the interpretational and inferential stages. The analysis in the section above indicates that the validity of categorical syllogisms can be checked using just two kinds of relation: inclusion ( $\sqsubset$ ) and exclusion ( $\sqsupset$ ). The effectiveness of Euler diagrams rests in the fact that they can make explicit the relations (inclusion and exclusion) implicit in categorical sentences, thereby aiding the avoidance of errors in the interpretational stage, and, more importantly, to replace the task of combining information in premises with a more concrete process of manipulating diagrams (cf., Mineshima, Sato, Takemura, & Okada, 2012).

Syllogism solving tasks can be compared using Venn and Euler diagrams depicting partial information, examining the difficulty of manipulating diagrams in reasoning processes. Deductive reasoning generally requires combining information regarding multiple premises. Such a task could naturally be replaced by a task involving a combination of presented diagrams. An inferential process of combining diagrams would be relatively easy to access, possibly triggering a construction of a diagrammatic procedure. Deductive reasoning tasks can be classified into two cases: validity checking and invalidity checking. The following section examines cases of validity and invalidity checking using Euler and Venn



diagrams. In addition, the relevant properties of diagrams that aid syllogistic reasoning tasks are discussed.

### Case I : Checking validity using diagrams

**(1) Solving syllogisms using EUL diagrams.** This case describes a procedure for checking the validity of a syllogism using Euler diagrams. As a concrete example, consider the case of the syllogism *All A are B, No C are B; therefore No C are A*. Suppose one is provided with two premise sentences and corresponding diagrams as external representations, as illustrated in the cognitive model in Fig. 2.2 of Section 1. The procedure is shown in Fig. 2.10. Here, the premise *All A*

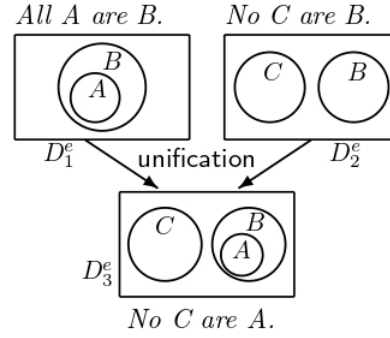


Fig. 2.10 Solving a syllogism *All A are B, No C are B; therefore No C are A* with Euler diagrams

*are B* is associated with  $D_1^e$ , where the relation  $A \sqsubset B$  holds, and the premise *No C are B* is associated with  $D_2^e$ , where the relation  $C \vdash B$  holds. These diagrams make explicit the relational information contained in the premise sentences. In the present setting, these two diagrams are presented externally to reasoners. In general, syllogistic reasoning with Euler diagrams involves a step of combining premise diagrams, referred to as the *unification* step. The operation of combining two diagrams  $D_1^e$  and  $D_2^e$  in Fig. 2.10 is an instance of an application of the unification rule.<sup>3</sup> In this case, the unification process involves identifying the circle  $B$  and ensuring that all the relations in the premise diagrams hold. The resulting diagram,  $D_3^e$ , has three relations:  $A \sqsubset B$ ,  $C \vdash B$  and  $A \vdash C$ . The first two are inherited from the premise diagrams  $D_1^e$  and  $D_2^e$ , and the last one (the exclusion

<sup>3</sup>The rule of unification plays a central role in formalizing reasoning with Euler diagrams. See Mineshima et al. (2012a).

relation  $A \vdash C$ ) is created as a by-product of the unification process. This new relation  $A \vdash C$  corresponds to the sentence *No C are A* (and *No A are C*), and hence, one can arrive at a valid conclusion for this syllogism.

An important characteristic of the unification process 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”. For the process of unifying diagrams, two rules determine the spatial relationship between objects in the conclusion diagram. Both are concerned with the natural properties of inclusion and exclusion relations. Namely, for any circle  $X$  and for any circle  $Y$  and  $Z$ ,

(C1) 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  $D_1 + D_2$ .

(C2) if  $X$  is inside  $Y$  in one diagram  $D_1$  and  $Y$  is outside  $Z$  in another diagram  $D_2$ ,  $X$  is outside  $Z$  in the combined diagram  $D_1 + D_2$ .

In the example in Fig. 2.10, the relation  $A \vdash C$  is obtained using (C2). Note that these two rules have counterparts in inference rules in **GS**: (C1) corresponds to the  $(\sqsubset)$  rule and (C2) corresponds to the  $(\vdash)$  rule. The constraints (C1) and (C2) appear to be so natural and intuitive that even users without explicit training in diagrammatic reasoning can exploit them to draw a correct conclusion without much effort.

Theoretically, the inference rules  $(\sqsubset)$  and  $(\vdash)$  which are crucial for deriving valid syllogisms, are simulated in terms of the spatial constraints, (C1) and (C2). Such a simulation can occur in actual syllogistic reasoning with external diagrams. For example, a procedure using the  $(\vdash)$  rule, enabling the derivation of  $A \vdash C$  from  $A \sqsubset B$  and  $B \vdash C$  (a linguistic case of Fig. 2.3 in Section 2.2), can be manifested by perceiving the spatial relationships between diagrammatic objects as seen in Fig. 2.10. Thus, it can be argued that sentential (linguistic) premises themselves do not provide untrained reasoners with specific procedures for solving syllogisms in terms of  $(\sqsubset)$  and  $(\vdash)$ , such as the ones we saw in the last section; in contrast, externally presented Euler diagrams provide reasoners with a concrete problem-solving procedure based on an intuitive understanding of constraints such as (C1) and (C2).

As a second example, consider a syllogism with existential sentences as a premise. An example, shown in Fig. 2.11, is the diagrammatic counterpart of the linguistic derivation in GS, shown in Fig. 2.4 in Section 2.2. Here the premises are *All B are A* and *Some C are B*. The premise *All B are A* is associated with

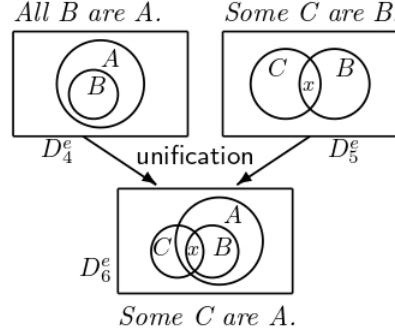


Fig. 2.11 Solving a syllogism *All B are A*, *Some C are B*; therefore *Some C are A* with Euler diagrams

the diagram  $D_4^e$ , where  $B \sqsubset A$  obtains; the premise *Some C are B* is associated with the diagram  $D_5^e$ , where  $x \sqsubset C, x \sqsubset B, B \bowtie C$  obtain. In this case, neither the constraint (C1) nor (C2) can be applied. This means that neither inclusion nor exclusion relations between circles  $A$  and  $C$ , i.e.,  $A \sqsubset C$ ,  $C \sqsubset A$ , and  $A \sqsupset C$ , are inferable from the information contained in the two premises. In such a case, the circles  $A$  and  $C$  must be placed in such a way that they partially overlap each other, so that  $A \bowtie C$  holds. Specifically, the relevant rule is the following. For any circles  $X$ ,  $Y$  and  $Z$ ,

- (C3) If none of the relations  $X \sqsubset Y$ ,  $Y \sqsubset X$ , or  $X \sqsupset Y$  holds in the combined diagram, put  $X$  and  $Y$  in such a way that  $X \bowtie Y$  holds.

In the case of the syllogism in Fig. 2.11, unifying the premise diagrams leads to the diagram  $D_6^e$ , where we have  $B \sqsubset A, \sqsubset C, x \sqsubset B, B \bowtie C, x \sqsubset A, A \bowtie C$ . Here the relation  $x \sqsubset A$  is newly introduced by the application of (C1) to  $x \sqsubset B$  and  $B \sqsubset A$  and  $A \bowtie C$  by the application of (C3). Then, by observing that the relations  $x \sqsubset A$  and  $x \sqsubset C$  hold on the unified diagram, one can draw the valid conclusion *Some C are A* (and *Some A are C*). Here, the process of constructing the unified diagram visually simulates the linguistic derivation shown in Fig. 2.4 in Section 2.2. In particular, the premise diagrams  $D_4^e$  and  $D_5^e$  make manifest the relational information that is implicit in the existential premise, and then the

process of combining the information is replaced by the process of unifying the diagrams by means of applications of the constraints (C1) and (C3).

The diagrammatic construction in Fig. 2.11 can be viewed as a manifestation of the free ride property of Euler diagrams, in that the relation  $A \bowtie C$  is *newly* introduced in the conclusion diagram and the reasoner can observe this relationship from it. However, an important difference between the case of the construction in Fig.2.10 and the case of Fig. 2.11 is that, in the former case, the fact that circle  $C$  is outside circle  $A$  ( $C \vdash A$ ) is forced by a natural spatial constraint (C2), whereas in the latter case, that fact that circle  $A$  and  $C$  partially overlap is caused by the syntactic convention introduced to handle partial information in Euler diagrams. The constraints (C1) and (C2) are both natural in that they are both solely based on the nature of spatial relationships about inclusion and exclusion relations, and other possibilities are not conceivable. Thus, these two constraints lead reasoners to determine the spatial relationship between circles almost automatically. In contrast, the constraint (C3) is conventional; thus, reasoners are forced to perform an operation where positioning circles are partially overlapped *consciously* in reasoning processes. Hence, the exploitation of (C3) seems to require more effort than that of (C1) and (C2).

**(2) Solving a syllogism using Venn diagrams.** An abstract solving process of the syllogism using Venn diagrams is illustrated in Fig. 2.12, where the premise *All B are A* is represented by  $D_1^v$ , and the premise *Some C are B* by  $D_2^v$  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  $D_1^v$ , and circle  $A$  to  $D_2^v$ . Then, by superposing the shaded region of  $D_3^v$  on  $D_4^v$ , one can obtain diagram  $D_5^v$ , 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. 2.10, 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. Comparing syllogistic inference GS and Venn diagrammatic inference reveals the same outcomes, even though the different events are clearly executed in two different inference systems.

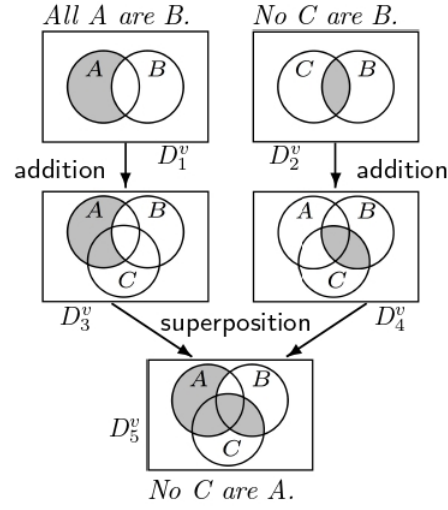


Fig. 2.12 Solving a syllogism *All A are B, No C are B; therefore No C are A* with Venn diagrams.

### Case II : Checking invalidity using diagrams

This section discusses the case of invalid inferences, which is considered to be particularly difficult for untrained reasoners (cf. Revlis, 1975; Khemlani & Jonhson-Laird, 2012) and hence deserves special attention. Consider a task in which two categorical sentences are given and the reasoner is required to judge that no valid conclusion can be drawn from them with respect to the relevant terms. Such tasks are referred to as *NVC tasks*. How do Euler diagrams help reasoners solve NVC tasks? As an illustration, consider a syllogism with the premises *All B are A* and *No C are B*. Fig. 2.13 shows the problem solving process suggested by the convention introduced in Section 3.1.

The premise *All B are A* is associated with a diagram  $D_7^v$  in which the relation  $B \sqsubset A$  holds, and the premise *No C are B* is associated with a diagram  $D_8^v$  in which the relation  $C \vdash B$  holds. By unifying these two diagrams, one can obtain the conclusion diagram  $D_9^v$ , where the relation  $A \bowtie C$  is newly introduced via the constraint (C3). The fact that  $A \bowtie C$  holds on the conclusion diagram indicates that no specific semantic information about the terms *A* and *C* can be drawn from the premises. This amounts to saying that there is no valid conclusion with respect to *A* and *C* (except trivial ones such as  $A \sqsubset A$ ) in this syllogism. Here again, we can see that Euler diagrams associated with sentential premises play a

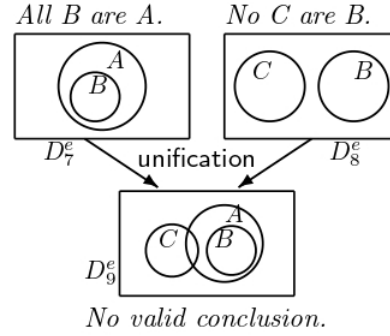


Fig. 2.13 Solving an NVC task using Euler diagrams

dual role in the process of checking the *invalidity* of a syllogism: first, they make explicit the relational information underlying categorical sentences; second, the unification of premise diagrams using the constraints (C1) and (C2) leads us to an understanding of what relational information can be obtained in a given inference; when no particular inclusion or exclusion relation is newly introduced by the unification, that is, when the situation is as described in (C3), the reasoner can conclude that there is no valid conclusion for the inference.

Note here that the conclusion diagram  $D_9^e$  not only explicitly delivers the *positive* information as in (1) and (2) below but also implicitly conveys the *negative* information as in (3).

- (1) The information  $B \sqsubset A$  follows from the premises  $D_7^e$  and  $D_8^e$ .
- (2) The information  $B \vdash C$  follows from the premises  $D_7^e$  and  $D_8^e$ .
- (3) Regarding the semantic relation between  $A$  and  $C$ , no meaningful information follows from  $D_7^e$  and  $D_8^e$ . That is, neither  $A \sqsubset C$ ,  $C \sqsubset A$  or  $A \vdash C$  are deducible from the premises  $D_7^e$  and  $D_8^e$ .

Thus, it can be concluded that there is no valid conclusion for this syllogism. Importantly, this shows that by employing a convention that enables us to express partial information, the process of checking invalidity can be conducted by a unification process in a similar way to the case of processes of checking validity.

To appreciate the nature of the information as presented in (3), it is crucial to distinguish between what a diagram delivers individually compared with what a diagram delivers in a *chain of inferences*. For instance, what the diagram  $D_9^e$  itself

delivers is that the set **A** is a subset of the set **B**, and that the set **B** is disjoint from the set **C**. If one considers  $D_9^e$  separately from other diagrams, the relation  $A \bowtie C$  is semantically vacuous, contributing nothing to the interpretation of this diagram. However, examining diagram  $D_3^e$  as a diagram appearing in the overall process in Fig. 2.13 reveals that *nothing follows from the two premises*, constituting a kind of *meta-information* about this argument. Importantly, a diagram in a logical inference can implicitly communicate *negative* information of the kind stated in (3): the diagram depicts exhaustive information obtained from the premises up to the current stage in the logical inference.

The procedure of checking the invalidity of inferences described here is distinct from the standard procedure, according to which an inference is judged to be valid if one can construct a counter-example where all the premises are true but the conclusion is false. Note that some existing proposals using diagrams are also based on such an idea of counter-example construction. Thus, in Lewis Carroll's conception of logic diagrams (Carroll, 1896), an inference is invalid if it is impossible to superpose all the premises and the *negation* of the conclusion. See Geach (1976) and Lear (1986) for discussion. The current diagrammatic procedure is distinctive in that it does not depend on a process of negating the conclusion: the information that there is no valid conclusion with respect to the two given terms can be obtained in a direct way, via a process of combining premises. It could be argued that such a procedure is based on the view that the fact not described in the representations does not hold. In this respect, the processes are similar to a well known formalization of commonsense reasoning, namely, circumscription (McCarthy, 1980, 1986), where the considerations are circumscribed in the described things and such a claim is explicitly provided as a inference rule in the logical system. Generally speaking, while such an efficient view could be explicit in reasoning using diagrammatic representations, it is not necessarily explicit in usual reasoning with linguistic representation.

## 2.4 Predictions

As we have seen, Euler diagrams seem to be relatively easy to handle even for those users who are not trained to manipulate them in syllogism solving. This can be explained by appealing to the fact that the relational structures in syllogistic (quantified) sentences matches with the inclusion-exclusion structures in Euler

diagrams; thanks to the correspondence between these two structures, syllogistic inferences (which require abstract and analytic modes of thinking and thus are not easy to conduct for untrained participants) can be replaced by concrete process of manipulations of diagrams. The essential steps involved in the manipulations of Euler diagrams are unification processes, i.e. those processes in which the inclusion and exclusion relations between objects in the unified diagrams are effectively determined using the constraints (C1), (C2), and (C3). Such a unification process consist of steps of matching an object (a circle or a point) with another object and then determining the diagrammatical relationships between the other objects. Given the fact that deductive reasoning generally requires combining the information in premises, these unification processes seem to be natural enough so that they would be immediately accessible to users. Thus I expect that users could exploit natural constraints of diagrams and extract the right rules to draw a conclusion from Euler diagrams themselves.

By contrast, Venn diagrams seem to be relatively difficult to handle in solving syllogisms. In order to unify 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, as indicated in Fig. 2.12. 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. I 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.

I will say that diagrammatic representations are *self-guiding* if the syntactic manipulation of diagrams are automatically triggered even for participants without explicit prior knowledge of inferential rules or strategies. Then, my hypothesis amounts to saying that in syllogistic reasoning tasks, Euler diagrams are self-guiding, whereas Venn diagrams are not.

Based on these considerations, I predict that (P1) the performance in syllogism solving would be better when participants use Euler diagrams than when they use Venn diagrams. As more specific predictions I also expect that (P2) the performance in syllogism solving with Euler diagrams would be better when the convention of crossing is involved in syllogistic tasks than when it is not, and that



(P3) the performance in syllogism solving would be better when participants use Euler diagrams than when they use Venn diagrams even in NVC tasks.

In order to clarify the Inferential efficacy of logic diagrams, we need to specify the interpretational efficacy, and then distinguish them. 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, of a sentential material (cf. Newstead & Griggs, 1983). It is expected that both Euler diagrams and Venn diagrams may help participants avoid such interpretational errors in linguistic syllogistic reasoning: conversion errors and figural effects.

**Conversion errors.** As stated in Section 1, the categorical sentence “All  $A$  are  $B$ ” is sometimes misinterpreted as equivalent to “All  $B$  are  $A$ .” Similarly, “Some  $A$  are not  $B$ ” is sometimes misinterpreted as equivalent to “Some  $B$  are not  $A$ ” (cf. the experimental results in Newstead & Griggs, 1983). As shown by the experiments on *illicit conversion error* in Chapman and Chapman (1959) and Dickstein (1981), such misinterpretations may cause errors in syllogism, in particular those that have no valid conclusions (AA2, AA3, AA4, AI2, AI4, IA1, IA2, AE1, AE3, EA3, EA4, AO1, AO3, AO4, OA1, OA2, and OA4 types in my experiment). For example, in the case of AE1 syllogism, the first premise “All  $B$  are  $A$ ” is often misinterpreted as equivalent to “All  $A$  are  $B$ ,” leading participants to select the invalid conclusion “No  $C$  are  $A$ .” Note that in Euler diagrams, “All  $A$  are  $B$ ” and “All  $B$  are  $A$ ” correspond to  $D_1^e$  and  $D_2^e$  of Fig. 2.14, respectively, and in Venn diagrams, they correspond to  $D_1^v$  and  $D_2^v$ , respectively. Here, one can immediately see that these two diagrams are topologically different, and hence, deliver different information. (Similarly for “Some  $A$  are not  $B$ ” and “Some  $B$  are not  $A$ ,” which are represented in Euler and Venn diagrams in the same way; see  $D_3$  and  $D_4$  below). Thus, the use of diagrams seems to block the errors caused by the misinterpretation of categorical sentences. This point would be supported by the results: the performance in the class of syllogisms mentioned above (i.e., those that have no valid conclusions) is better when Euler and Venn diagrams are available to participants.

**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  $A$ ” (cf. New-

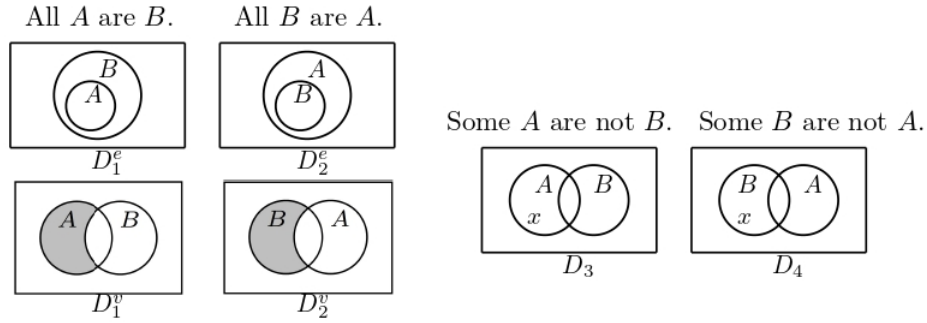


Fig. 2.14 Topologically non-identical pairs in Euler and Venn diagrams

stead & Griggs, 1983). 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 sentences are represented as shown in Fig. 2.15. 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.

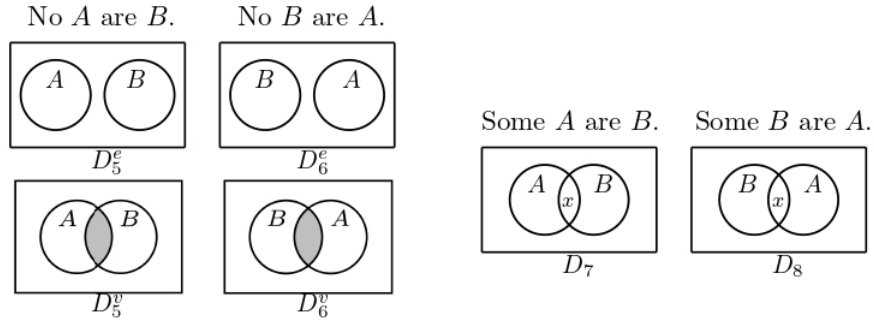


Fig. 2.15 Topologically identical pairs in Euler and Venn diagrams

## Chapter 3

# Empirical studies of the inferential efficacy

Based on the considerations described in earlier chapters, the current chapter presents an empirical study of efficacy of logic diagrams. Section 1 describes an experiment comparing participants' performance in syllogism solving using Euler diagrams and Venn diagrams. Section 2 examines whether the pattern of results in Experiment 1 also emerges when a linear variant of Euler diagrams, in which set-relationships are represented by one-dimensional lines, is used. Section 3 describes an fMRI experiment comparing hemodynamic changes in reasoners with and without Euler diagrams. In Section 4, children's performance is compared in syllogistic reasoning tasks externally supported by Euler and Venn diagrams. The results of the four experiments provide evidence supporting the hypothesis that Euler diagrams have inferential efficacy, that is, they are effective in inferential processes of combining the premise information.

### 3.1 Reasoning with Euler and Venn diagrams: Experiment 1

In this section, I examine the predictions of inferential efficacy (P1, P2, and P3) and interpretational efficacy (conversion error and figural effect) described in Section 4 of Chapter 2, focusing on a typical version of relation-based diagrams, i.e., Euler diagrams. An earlier version of this work appeared in Sato, Mineshima, and Takemura (2010a).

### 3.1.1 Method

In order to test my hypothesis, I provided the participants in my experiment only with instructions on the meanings of diagrams and required them to solve reasoning tasks without any instruction on how to manipulate diagrams in syllogism solving. I first conducted a pretest to check whether participants understood the instructions correctly. The pretest was designed mainly to see whether participants correctly understand the conventional devices of each system, in particular, the convention of crossing in the Euler and Venn systems and shading in the Venn system. I then 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.

#### Participants

Two hundred and thirty-six undergraduates (mean age  $20.13 \pm 2.99$  SD) in five introductory 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.<sup>1</sup> The participants were divided into three groups: the Linguistic group, the Euler group, and the Venn group. The Linguistic group consisted of 66 students. Of them, I excluded 21 students: those who left the last more than three questions unanswered (19 students) and those who had participated in my pilot experiments conducted before (2 students). The Euler group consists of 68 students. Of them, I excluded 5 students: those who left the last more than three questions unanswered (3 students) and those who had participated in my pilot experiments conducted before (2 students). The Venn group consists of 102 students. Of them, I excluded 34 students: those who left the last more than three questions unanswered (27 students) and those who had participated in my pilot experiments conducted before (7 students).

---

<sup>1</sup>I 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 I 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 quantifier corresponding to *No*, I 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 thesis.

## 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 Fig. 3.1 and Fig. 3.2, respectively.

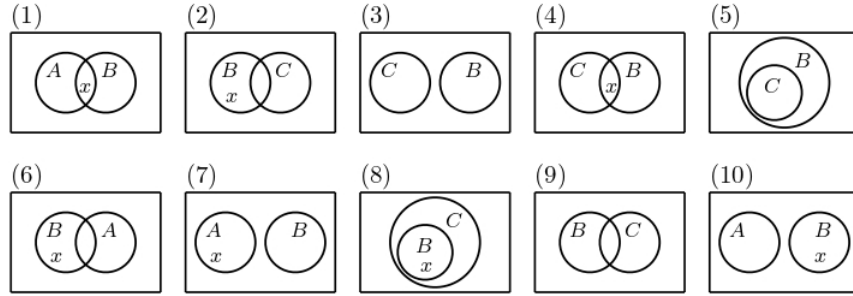


Fig. 3.1 Euler diagrams used in the pretest

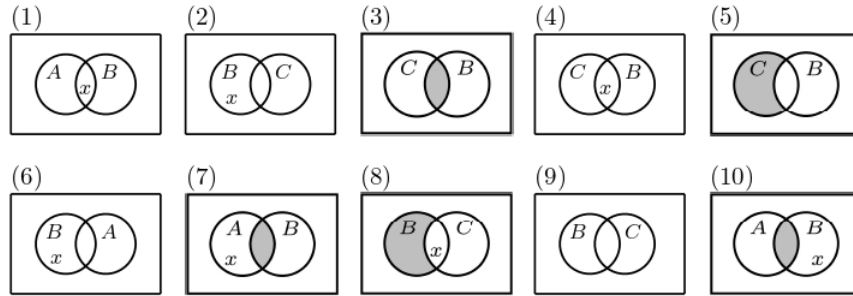


Fig. 3.2 Venn diagrams used in the pretest

The participants were asked to choose, from a list of five possibilities, the sentences corresponding to a given diagram. Examples are given in Figs. 3.3 and 3.4. 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. Their cutoff point was set to be eight. These cutoff points were

chosen carefully, based upon the results of my pilot experiments. The total time in Euler and Venn groups was 5 minutes. Before the pretest, the participants were presented with the examples in Figs. 3.3 and 3.4.

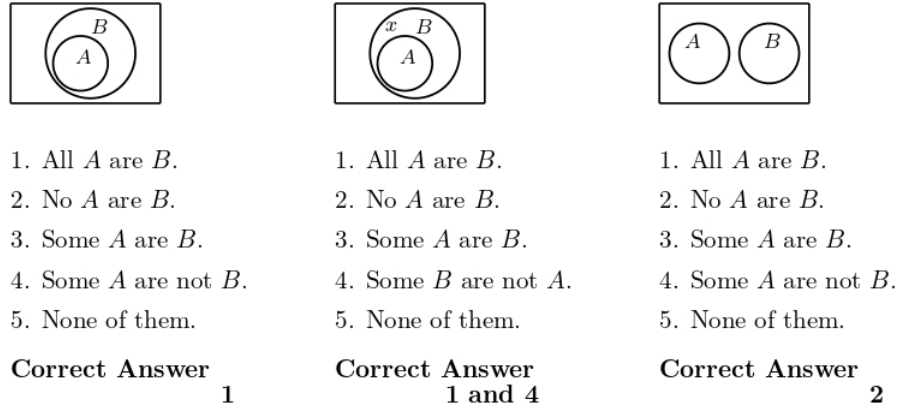


Fig. 3.3 The examples in the pretest of the Euler diagrams

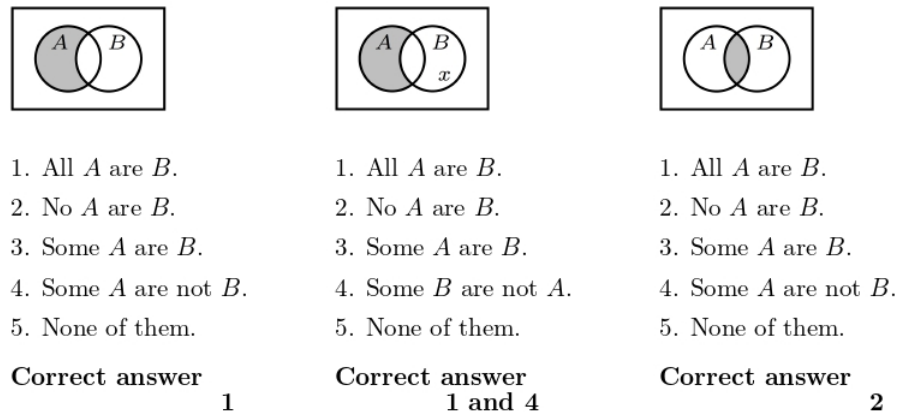
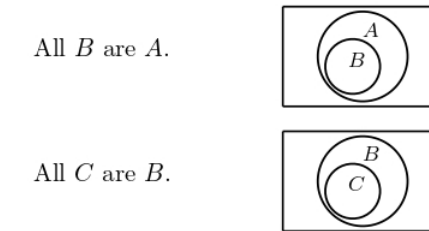


Fig. 3.4 The examples in the pretest of the Venn diagrams

**Syllogistic reasoning tasks.** The participants in the Euler group were given syllogisms with Euler diagrams (such as the one in Fig. 3.5). The participants in the Venn group were given syllogisms with Venn diagrams (such as the one in Fig. 3.6), and participants in the Linguistic group were given syllogisms without diagrams. The participants were presented with two premises and were asked to

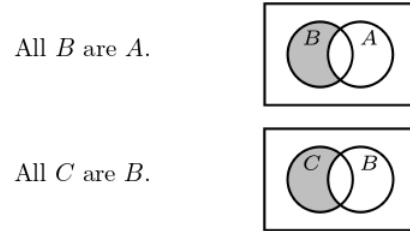
choose, from a list of five possibilities, a sentence corresponding to the correct conclusion. The list consists of *All*-, *No*-, *Some*-, *Some-not*, and *NoValid*. The subject-predicate order of each conclusion was  $CA$ . I gave 31 syllogisms in total, out of which 14 syllogisms had a valid conclusion 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-minute power test, and each task was presented in random order (10 patterns were prepared). Before the test, the example in Fig. 3.5 was presented to participants in the Euler group, and the one in Fig. 3.6 to participants in the Venn group.



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.3.5 An example of syllogistic reasoning task of the Euler group



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.3.6 An example of syllogistic reasoning task of the Venn group

## Procedure

All three groups were first given 1 minute 30 seconds to read one page of instructions on the meaning of categorical statements (the instructions are given in Appendix B.1.1). In addition, the Euler group was given 2 minutes to read two pages of instructions on the meaning of Euler diagrams (the instructions are given in Appendix B.1.2), and the Venn group was given 2 minutes to read two pages of instructions on the meaning of Venn diagrams (the instructions are given in Appendix B.1.3). Before the pretest, the Euler and Venn groups were given 1 minute 30 seconds to read two pages of instructions on the pretest. Finally, before

the syllogistic reasoning test, all three groups were given 1 minute 30 seconds 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 my pilot experiments.

### 3.1.2 Results

#### Pretest

The accuracy rate of each item in the pretest of Euler diagrams, listed in Fig. 8 was (1) 77.8%, (2) 77.8%, (3) 90.5%, (4) 81.0%, (5) 69.8%, (6) 84.1%, (7) 49.2%, (8) 58.7%, (9) 79.4%, and (10) 82.5%, respectively. One major source of error was the misunderstanding of the convention of crossing; participants tended to incorrectly select both *Some-* and *Some-not* in (1), (2), (4), (5), (6), and (9). This error was observed in 18 students (out of 63 students), who scored less than 8 on the pretest (out of 12). In the following analysis, I exclude these 18 students and refer to the other 45 students as the Euler group. The correlation coefficient between the scores of pretest and of syllogistic reasoning tasks in the total Euler group ( $N = 63$ ) was substantially positive with 0.606.

The accuracy rate of each item in the pretest of Venn diagrams, listed in Fig. 9 was (1) 85.2%, (2) 75.0%, (3) 74.0%, (4) 82.0%, (5) 22.1%, (6) 66.0%, (7) 35.3%, (8) 35.3%, (9) 77.9%, and (10) 66.1%, respectively. One major source of error was also the misunderstanding of crossing (partially overlapping circles); participants tended to incorrectly select both *Some-* and *Some-not* in (1), (2), (4), (5), (6), and (9). This error was observed in 38 students (out of 68 students) who scored less than 8 on the pretest (out of 12). In the following analysis, I exclude these 38 students and refer to the other 30 students as the Venn group. The correlation coefficient between the scores of pretest and of syllogistic reasoning tasks in the total Venn group ( $N = 68$ ) was substantially positive with 0.565.

#### Syllogistic reasoning tasks

Fig. 3.7 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 66.5%, and the rate for the Euler group was 85.2%. The results of each syllogistic type are shown in Table B.1 of Appendix B.2.

These data were subjected to a one-way Analysis of Variance (ANOVA). There



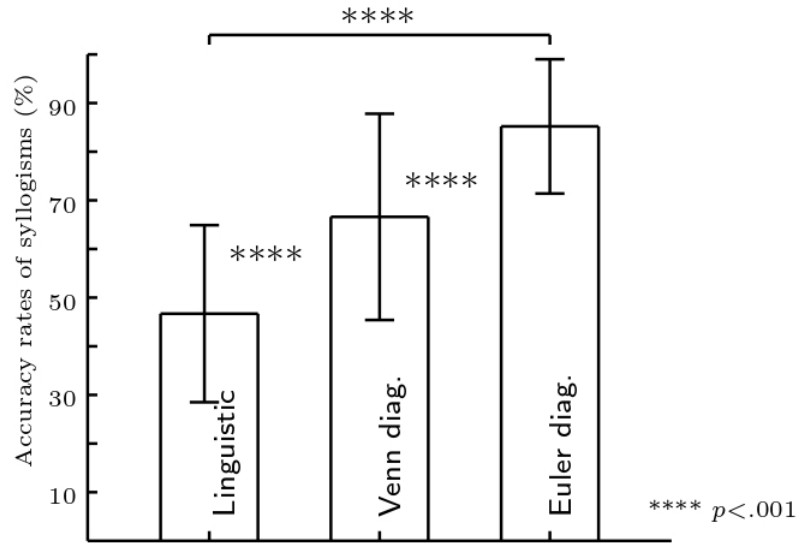


Fig. 3.7 The average accuracy rates of 31 total syllogisms in the Linguistic group, the Venn group, and the Euler group (error-bar refers to SD)

was a significant main effect,  $F(2, 117) = 52.515, p < .001$ . 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 85.2% for the Euler group ( $F(1, 88) = 10.247, p < .001$ ). (ii) The accuracy rate of reasoning tasks in the Euler group was higher than that in the Venn diagrammatic group: 66.5% for the Venn group and 85.2% for the Euler group ( $F(1, 73) = 4.421, p < .001$ ). (iii) The accuracy rate of reasoning tasks in the Venn group was higher than that in the Linguistic group: 66.5% for the Venn group and 46.7% for the Linguistic group ( $F(1, 73) = 4.744, p < .001$ ). It should be noted that if we include those participants who failed the pretest, we still obtain similar results in each comparison: for (i) and (ii), there were significant differences,  $p < .001$ ; for (iii), there were no significant differences. The rate for the total Euler group (including those who failed the pretest) was 77.8% and that for the total Venn group (including those who failed the pretest) was 52.9%.

Detailed performance data of the Euler group are provided in Table 3.1. In this table, "no-crossing" refers to syllogisms not involving the convention of crossing, and "crossing" refers to those involving the convention of crossing. 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 subject-predicate order *CA* in conclusions).

Table 3.1 The average accuracy rates of the three cases in the Euler group (compared with the Venn group)

	Valid Conclusion		NVC
	<i>no-crossing</i>	<i>crossing</i>	
Euler group	97.2%	81.5%	85.7%
Venn group	88.3%	69.3%	59.5%

These data were subjected to a two-way ANOVA. There was a significant main effect for the difference between Euler and Venn diagrams, ( $F(1, 73) = 22.593$ ,  $p < .001$ ). There was a significant main effect for the difference between syllogisms involving crossing and those which do not, ( $F(2, 146) = 23.381$ ,  $p < .001$ ). There was a significant interaction effect for two factors, ( $F(2, 146) = 4.135$ ,  $p < .05$ ). Multiple comparison tests were conducted by Ryan's procedure.

(i) In the Euler group, the performances of the valid syllogisms not involving crossing were better than those of the valid syllogisms involving crossing ( $F(1, 88) = 3.880$ ,  $p < .05$ ), and than those of the NVC syllogisms involving crossing ( $F(1, 88) = 2.850$ ,  $p < .05$ ).

(ii) The performances of the valid syllogisms not involving crossing in Euler group were better than those in Venn group at the reduced threshold of  $p < .10$  ( $F(1, 219) = 3.213$ ). The performances of the valid syllogisms involving crossing in Euler group were significantly better than those in Venn group,  $F(1, 219) = 6.070$ ,  $p < .05$ . The performances of the NVC syllogisms involving crossing in Euler group were significantly better than those in Venn group,  $F(1, 219) = 27.897$ ,  $p < .001$ .

A significant difference in performance was found in the Linguistic group and the diagrammatic groups (Euler and Venn groups). This difference can be ascribed to a well-known interpretational bias in linguistic syllogistic reasoning: conversion errors and figural effects. The present results indicate that the two type of effects were blocked in both Euler and Venn diagrammatic groups.

**Conversion errors.** In the 17 syllogisms having no valid conclusion, as Table 3.2 indicates, 58.8% of the participants in the Linguistic group selected the illicit converted conclusion while the rate reduced to 9.1% in the Euler group and 30.2% in the Venn group. These data were also subjected to a one-way ANOVA. There

was significant main effect,  $F(2, 117) = 39.670, p < .001$ . Multiple comparison tests by Ryan's procedure of main effect yield the following results: (i) There was significant difference between the Linguistic group and the Euler group,  $F(1, 88) = 8.881, p < .001$ . (ii) There was significant difference between the Venn group and the Linguistic group,  $F(1, 73) = 4.578, p < .001$ . (iii) There was significant difference between the Euler group and the Venn group,  $F(1, 73) = 3.365, p < .005$ .

Table 3.2 The average error rates of conversion errors in the Linguistic, Euler and Venn group; a number in parentheses refers to the rates when we include those participants who failed the pretest

	Linguistic	Euler	Venn
Conversion error	58.8%	9.1%(16.3%)	30.2%(45.3%)

It should be noted that if I include those participants who failed the pretest, we still obtain similar results in each comparison: for (i) and (iii), there were significant differences,  $p < .001$ ; for (ii), there was significant differences,  $p < .05$ . The rate for the total Euler group (including those who failed the pretest) was 16.3% and that for the total Venn group (including those who failed the pretest) was 45.3%.

**Figural effects.** In fact, comparing EI1O and EI4O syllogisms, as Table 3.3 indicates, there was significant difference between EI1O  $no(B, A); some(C, B) : some-not(C, A)$  and EI4O  $no(A, B); some(B, C) : some-not(C, A)$  in the Linguistic group (62.2% for EI1O and 35.6% for EI4O) ( $t(44) = 11.000, p < .005$ . t-test, within-subjects design). In contrast, there was no significant difference between EI1O and EI4O in the Euler group (84.4% for EI1O and 75.6% for EI4O) ( $t(44) = 1.622, p = .100$ ). Further, there was no significant difference between EI1O and EI4O in the Venn group (70.0% for EI1O and 73.3% for EI4O) ( $t(29) = 1.000, p = .100$ ).

It should be noted that if we include those participants who failed the pretest, we still obtain similar results: there was no significant differences between EI1O and EI4O in each the total Euler group and the total Venn group,  $p = .10$ . The rates for the total Euler group (including those who failed the pretest) was 76.2% for EI1O and 66.7% for EI4O and that for the total Venn group (including those who failed the pretest) was 57.4% for EI1O and 54.4% for EI4O.

Table 3.3 The average accuracy rates of syllogisms of EI1 and EI4 types in the Linguistic, Euler and Venn group; a number in parentheses refers to the rates when we include those participants who failed the pretest

<i>figure</i>	<i>1st &amp; 2nd premises</i>	Linguistic	Euler	Venn
EI1O	<i>no(B, A); some(C, B)</i>	62.2%	84.4% (76.2%)	70.0% (57.4%)
EI4O	<i>no(A, B); some(B, C)</i>	35.6%	75.6% (66.7%)	73.3% (54.4%)

### 3.1.3 Discussion

Syllogistic reasoning performance in the Euler groups was significantly better than in the Venn groups. This result supports my prediction of inferential efficacy (P1) and corroborates the role of Euler diagrams in reasoning processes. As expected in (P2) the performance in syllogism solving with Euler diagrams was significantly better when the convention of crossing was involved in syllogistic tasks compared with when it was not. This result indicates that the exploitation of the convention of crossing requires more effort in reasoning processes. In addition, as expected in (P3) syllogism solving performance was significantly better when participants used Euler diagrams compared with when they used Venn diagrams, even in NVC tasks. Studies in the psychology of deductive reasoning have established that NVC tasks are difficult to solve (cf. Revlis, 1975). However, the current results indicate that Euler diagrams are effective in aiding NVC task performance.

The performance of syllogistic reasoning in the Euler and Venn groups was significantly better than in the Linguistic group. These results support the notion that Euler diagrams help participants solve syllogisms. However, 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. This difference in training may have had a major impact on their differences in the performance in syllogism solving.

However, this potential confound could be avoided by comparing between the Euler group and the Venn group, since both received substantial instruction about the categorical sentences, and underwent a similar number of practice trials. This comparison revealed that the performance of the Euler group was significantly better than that of the Venn group, suggesting that the results were not caused by practice effects alone. The difference in performance between these two groups

may be due to Euler diagrams not only contributing to the correct interpretations of categorical sentences but also playing a substantial role in the inferential processes of syllogism solving. Euler diagrams themselves could aid participants in inferential processes and thereby solving syllogistic reasoning tasks. Given the current experimental design, in which participants were not taught strategies for combining diagrams in syllogism solving, the empirical findings support the hypothesis that Euler diagrams are distinctive in that they are *self-guiding*, as discussed in Section 4 of Chapter 2.

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. 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. Supplementary analysis of participants who failed the pretest indicated that the difference between the Linguistic group and the Venn group was not caused by the exclusion of participants with poor performance, but from the interpretational effect of Venn diagrams. On the other hand, a figural effect in the performance between EI1O and EI4O syllogisms was observed in the Linguistic group but not in the Venn group. The supplementary analysis including participants who failed the pretest revealed similar results. As expected from the prediction of the interpretational effect (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.

The results with Euler diagrams described above are partially consistent with the findings of Ford (1994), who investigated cognitive strategies of syllogistic reasoning by analyzing participants' verbal and written protocols. Interestingly, it was reported that participants' strategies were generally divided into two types: (i) verbal strategies using the substitution method, and (ii) spatial strategies using Euler diagrams. It should be noted here that the Euler diagrams constructed by the spatial reasoners were similar to externally presented EUL diagrams, in that a single diagram can represent a categorical sentence, as mentioned in Section 3.1 of Chapter 2. Strictly speaking, the spatial reasoners did not depict the

points to express the existence of the objects. However, based on an analysis of participants' complementary verbal explanations of the written protocols, Ford claims that the spatial reasoners regarded the intersection region in the diagram corresponding to *Some A are B* as a non-empty set.

Given the correspondence, diagrams that were self-constructed by the spatial reasoners would also be expected to have interpretational effects. The following findings of Ford's study appear to be consistent with the current results regarding the figural effect, extending the interpretational effect of externally given diagrams to the effect of self-constructed diagrams: (i) There was no substantial difference between EI1O and EI4O syllogisms in spatial reasoners (40.0% for EI1O and 46.7% for EI4O). (ii) In contrast, there was a substantial difference between EI1O and EI4O syllogisms in verbal reasoners (73.3% for EI1O and 13.3% for EI4O). (However, it should be noted that Ford did not conduct a statistical analysis for this comparison).

Furthermore, these self-constructed diagrams appear would also be expected to have inferential effects, since the diagrams could avoid combinatorial explosion. The results of Ford's experiment indicate that the spatial reasoners unified premise diagrams for many syllogistic types. Furthermore, it was reported that when the spatial constraints allowed only one way to unify the premise diagrams (in the case of a valid syllogism with *All* premises and both *All* and *No* premises), performance was better than in situations with two possible combinations of diagrams (in the case of a valid syllogism with both *All* and *Some* premises), although no statistically significant differences were found. This result appears to be consistent with P2 in the current experiment, and provides initial support for the concrete manipulation of self-constructed Euler diagrams in syllogistic reasoning.

Importantly, the reasoners did not necessarily employ concrete manipulation of Euler diagrams during self-construction for all syllogistic types. In fact, when there were three possible combinations of premise diagrams under the spatial constraints (syllogisms with both *No* and *Some* or both *All* and *Some-not* premises), many participants did not select the most general combination. Performance in such cases was significantly poorer than in the other cases (less than 50%). This pattern of results was replicated in Bacon, Handly, and Newstead (2003), testing a greater number of participants in similar protocol experiments. The same observation applies to the NVC cases, although both Ford and Bacon et al. failed to

include these syllogistic types in their experiments. Thus, self-constructed Euler diagrams appear have only a partial inferential effect. Several of the observations discussed above imply a certain degree of effortlessness for EUL diagrammatic representation and its manipulation for syllogistic reasoning. That is, even if non-defective representational systems (including the convention on crossing circles) are not presented as external diagrams, novice reasoners can come up with an idea of EUL-like diagrams and, in some limited cases, manipulate the diagrams.

## 3.2 Reasoning with Linear diagrams: Experiment 2

The key assumption in Experiment 1 is that the form of diagrams mirrors the semantic information required in a given reasoning task, particularly the relational information in the case of syllogisms. Based on this assumption, any diagram that can make explicit the relational information encoded in a categorical sentence would be effective in supporting syllogistic reasoning. Experiment 2 aims to investigate whether this is true for a linear variant of Euler diagrams, where set-relationships are represented by one-dimensional lines, rather than by circles in a plane. A brief outline of Experiment 2 was shown in our previous work of Sato and Mineshima (2012). I hypothesize that linear diagrams would also provide an effective way of representing relational structures and reasoning about them. These findings may provide further support for my earlier hypothesis regarding inferential efficacy. Among various linear systems, a version of linear diagrams based on our EUL diagrams is used, as shown in Fig. 3.8.<sup>2</sup> Here, the only difference between EUL diagrams and linear diagrams is that each circle is replaced with a line. Thus, *All A are B* and *No A are B* can be clearly represented by the inclusion and the exclusion relations between lines. In addition, the conventional device of partially overlapping lines (corresponding to crossing circles in EUL diagrams) can be used with named points to represent existential sentences.

---

<sup>2</sup>Linear diagrams for categorical syllogisms were introduced by Leibniz (1903/1988) in the 17th century. Linear diagrams were developed by Lambert in 1764. More recently, Englebrechtsen (1992) provided a logical system of deductive inference with linear diagrams. Although it is known that linear diagrams have limited expressive power (cf. Lemon & Pratt, 1998), they are expressive enough to represent basic categorical syllogisms (cf. Politzer, van der Henst, Luche, & Noveck, 2006).

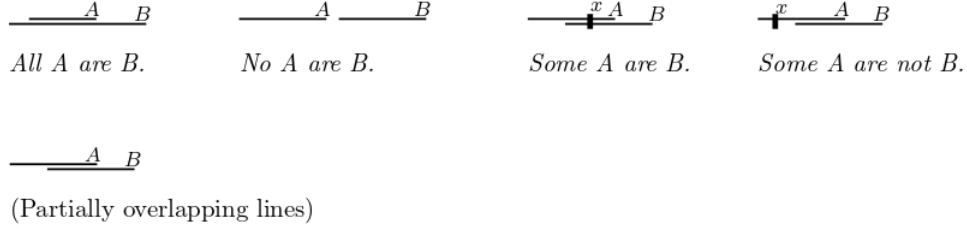


Fig.3.8 Representations of categorical sentences in Linear diagrams

A similar situation applies for solving processes of syllogisms. As an example, consider the case of a syllogism *All A are B, No C are B; therefore, No C are A*. This process is shown in Fig. 3.9, corresponding to the Euler diagrams shown in Fig. 2.10. By unifying the premises of two linear diagrams, the reasoner can obtain the desired information, based on the natural constraints of inclusion and exclusion relations. The case of NVC tasks in linear diagrams is also similar to the process described in Fig. 2.13 for EUL diagrams. The indeterminacy of the relationship between lines can be expressed according to the crossing convention. If such linear diagrams function as effectively as Euler diagrams, this would suggest that the effectiveness of external diagrams in syllogistic reasoning is not due to the particular shape (i.e., circles) of Euler diagrams.

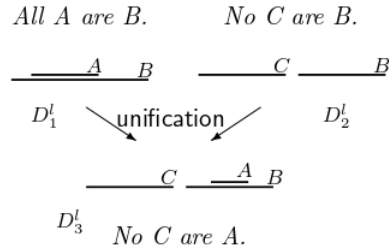


Fig. 3.9 Solving a syllogism *All A are B, No C are B; therefore No C are A* with Linear diagrams

The current study involves three key predictions: (P1') performance in syllogism solving would be better when participants used linear diagrams compared with when they used Venn diagrams; (P2') performance in syllogism solving with linear diagrams would be better when the crossing convention was involved in syllogistic tasks compared with when it was not; and (P3') performance in syllogism solving would be better when participants use linear diagrams than when



they use Venn diagrams even in NVC tasks. In addition, no significant difference would be expected between in syllogism solving performance with linear diagrams and Euler diagrams.

### 3.2.1 Method

The semantics of the linear diagrams used is essentially the same as that of Euler and Venn diagrams in Experiment 1. The experiment was conducted in the same manner as that of Experiment 1; the only difference is that in syllogistic reasoning tasks, Euler and Venn diagrams associated with premises are replaced by the corresponding linear diagrams.

#### Participants

Thirty-three undergraduates (mean age  $22.72 \pm 8.72$  SD) in an elementary philosophy class participated in the experiment. we called this group “the Linear group”. Of 33 students, I excluded 5 students: those who left the last more than three questions unanswered (4 students) and those who did not follow my instructions (1 student).

#### Materials

**Pretest.** The participants of the Linear group were presented with 10 diagrams listed in Fig. 3.10. The participants were asked to choose the sentences corresponding to a given diagram. Before the pretest, the participants were presented with the examples in Fig. 3.11.

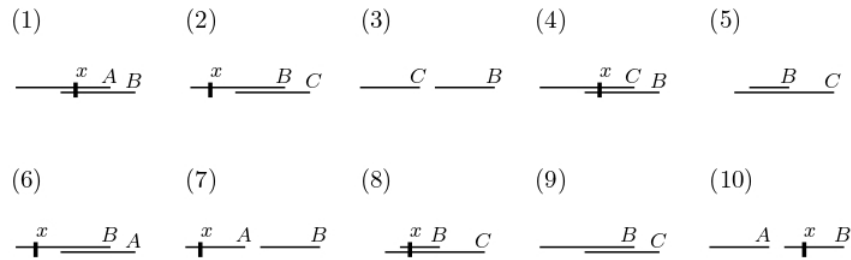


Fig. 3.10 Linear diagrams used in the pretest

**Syllogistic reasoning tasks.** The participants in the Linear group were given syllogisms with Linear diagrams (such as the one in Fig. 3.12). The partic-

<u>—</u> <u>A</u> B	<u>—</u> <u>A</u> <sup>x</sup> B	<u>—</u> A <u>—</u> B
1. All A are B.	1. All A are B.	1. All A are B.
2. No A are B.	2. No A are B.	2. No A are B.
3. Some A are B.	3. Some A are B.	3. Some A are B.
4. Some A are not B.	4. Some B are not A.	4. Some A are not B.
5. None of them.	5. None of them.	5. None of them.
<b>Correct Answer</b> 1	<b>Correct Answer</b> 1 and 4	<b>Correct Answer</b> 2

Fig.3.11 The examples in the pretest of the Linear diagrams

ipants were presented with two premises and were asked to choose a sentence corresponding to the correct conclusion.

## Procedure

The experiment was conducted in the same manner as Experiment 1. The instructions of sentences and diagrams (cf. Appendix B.1.4) were provided, pretests were conducted, and then the reasoning task were imposed.

All B are A.	<u>—</u> <u>B</u> A
All C are B.	<u>—</u> <u>C</u> 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.	
<b>Correct answer: 1</b>	

Fig.3.12 An example of syllogistic reasoning task of the Linear group

## 3.2.2 Results and Discussion

### Pretest

The accuracy rate of each item in the pretest of linear diagrams, listed in Fig. 3.10, was (1) 92.9%, (2) 85.7%, (3) 96.0%, (4) 96.0%, (5) 75.0%, (6) 86.0%, (7) 46.0%, (8) 61.0%, (9) 82.1%, and (10) 64.3%, respectively. 7 students scored less than 8 on the pretest (out of 12). In the following analysis, I exclude these 7 students and refer to the other 21 students as the Linear group. The correlation coefficient between the scores of pretest and of syllogistic reasoning tasks in the total Linear group ( $N = 28$ ) was substantially positive with 0.782.

### Syllogistic reasoning tasks

The average accuracy rate for the 31 syllogisms in the Linear group was 80.7%. The results of each syllogistic type are shown in Table B.2 of Appendix B.1. The Linear group data were compared with the Linguistic, Venn, and Euler groups in Experiment 1 using one-way analysis of variance (ANOVA). The results revealed a significant main effect,  $F(3, 140) = 37.734, p < .001$ . 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 80.7% for the Linear group ( $F(1, 64) = 7.112, p < .001$ ). (ii) The accuracy rate of reasoning tasks in the Linear group was higher than in the Venn diagram group: 66.5% for the Venn group and 80.7% for the Linear group ( $F(1, 49) = 2.741, p < .05$ ). (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 85.2% for the Euler group ( $F(1, 74) = 0.925, p = .10$ ). It should be noted that even when participants who failed the pretest were included, similar results were obtained in each comparison: for (i) and (ii), there were significant differences,  $p < .001$ ; for (iii), there were no significant differences. The accuracy rate for the total Linear group (including those who failed the pretest) was 71.1%. The performance of syllogistic reasoning in the Linear groups was significantly better than that in the Venn groups. This result supports my prediction of the inferential efficacy of the Linear method (P1') and indicates that there was no essential difference between linear diagrams and Euler diagrams.

In the Linear group, the average accuracy rate for valid syllogisms not involv-

ing crossing was 96.4%, the accuracy rate for valid syllogisms involving crossing was 76.6%, and the accuracy rate for NVC syllogisms involving crossing was 80.9%. These data were subjected to a one-way ANOVA. There was a significant main effect,  $F(3, 62) = 6.670, p < .001$ . Multiple comparison tests conducted using Ryan's procedure yielded the following results: performance for valid syllogisms not involving crossing were significantly better than for valid syllogisms involving crossing ( $F(1, 21) = 3.472, p < .005$ ), and for NVC syllogisms involving crossing ( $F(1, 21) = 2.717, p < .05$ ). As predicted by (P2'), syllogism solving performance with linear diagrams was significantly better when the convention of crossing was involved in syllogistic tasks, compared with when it is not. Performance with NVC syllogisms in the Linear group was significantly better than in the Venn group: 80.9% for the Linear group and 59.5% for the Venn group ( $F(1, 50) = 6.321, p < .05$ ). As predicted by (P3'), syllogism solving performance was significantly better when participants used linear diagrams compared with Venn diagrams, even in NVC tasks.

These results clearly supported the prediction that linear diagrams function as effectively as Euler diagrams in syllogistic reasoning. In turn, this finding provides evidence that the efficacy of external diagrams in syllogistic reasoning depends on the diagrams making explicit the semantic relations (such as inclusion and exclusion relations) in a way that is suitable for syntactic manipulation.

### 3.3 fMRI analysis of the efficacy of Euler diagrams: Experiment 3

The efficacy of external representation (in particular, diagrams presented externally) in several types of problem solving has been investigated in the literature of cognitive science and artificial intelligence (Glasgow, Narayanan & Chandrasekaran, 1995). The studies collected in the above reference have ranged widely over geometric theorem proving, solving physics problems such as pulleys systems, statistical thinking from bar and line graphs, and logical reasoning with Euler and Venn diagrams. As stated in Section 1 of Chapter 2, however, whether or not external diagrams play an essential role in reasoning remains controversial. For example, a seminal study by Larkin and Simon (1987) revealed that diagrammatic representations by direct exploitation of spatial and location information could be effective in the processes of search and recognition of information but

could not affect the inferential processes. Their claims could be summarized as the view that inference is solely dependent on semantic information of representation presented externally and is largely independent of ways of representing information. In contrast, Zhang and Norman (1994) reported that physical and spatial structures of external representations could constrain possible participant operations and provide an effective strategy for the Tower of Hanoi tasks. Their findings underlie the view that external representations can change the nature of tasks and reveal that external representations are effective in the dynamic inferential processes. An educational importance of the dynamic external representation has been discussed in the research that is referred to as “animated diagrams” (e.g. Lowe & Schnotz, 2008). On the basis of their findings, the efficacy of external diagrams can be divided into off-loading in interpretational processes and off-loading in inferential processes. However, research on external representation often fails to grasp the relationship between these two types. To discuss this problem, we focused on deductive reasoning, in which the processes of interpretation and inference have been studied for a long time (see also Chapter 1 for the related background).

The notion of “cognitive load” has been extensively utilized in the literature of cognitive neuroscience (e.g., Jonides et al., 1997; van den Heuvel et al., 2003). They designed the load by subtracting simple processes from complex ones, and on the basis this framework, they addressed the issue of neural basis involving task performances. Regarding deductive inference, Monti, Osherson, Martinez and Parsons (2007) compared brain activation during complex and simple inferences. In their design, the sentences used in their reasoning tasks were controlled to present participants with the same amount of syntactical complexity; however, the number of operations involving sentence integration was expected to be different. In reasoning stage rather than reading stage, they reported activation in the left rostral prefrontal cortex (PFC) (frontal pole; BA 10), peaking at  $[-36, 56, 8]$ , and bilateral medial PFC (BA 8), peaking at  $[-2, 28, 38]$ , which are associated with abstract information processing and executive control.

These findings on reasoning in the general context of neuroimaging studies can be applied to the particular examination on the efficacy of diagrams in reasoning. In my experimental situation, the usual reasoning tasks comprising sentences were compared with the reasoning tasks comprising both sentences and diagrams. In this situation, if the use of diagrams provides changes in activation of the frontal

gyrus including the rostral PFC (BA 10), it could count as an evidence that concrete processes of diagram manipulation substitute for abstract inferential processes. This would suggest that diagrams have the inferential efficacy, in which diagrams are syntactic objects that can be manipulated in certain ways; inferences could be made about a diagram itself. In contrast, if the use of diagrams does not provide any changes of activation of BA 10, it could count as an evidence that the efficacy of diagrams is restricted to interpretational processes, and diagrams are not efficient in inference. This would suggest that diagrams do not have the inferential efficacy.

In view of this application, I consider that the regions in which the activations in the condition without diagrams are higher than those in the condition including diagrams could provide a suggestion on the role which the diagrams would play in reasoning tasks. On the other hand, the regions in which there are no significant difference between the two conditions could give a suggestion on the modality-independent reasoning or the fundamental and primitive properties of human reasoning. Given the discussions of the relationship between spatial processing and deductive reasoning in the several review articles (Goel, 2007; Knauff, 2007; Prado, Chadha, & Booth, 2011), the findings of common activated regions in the two task situations would also be useful.

Here Experiment 3 will perform functional magnetic resonance imaging (fMRI) and compare performances of participants and hemodynamic changes in syllogism solving in cases where diagrammatic representations were used with those in cases where they were not used. This study is based on the work of Sato, Masuda, Someya, Tsujii, and Watanabe (in preparation).

### 3.3.1 Methods

#### Participants

Right-handed healthy volunteers ( $N = 32$ , mean age  $\pm$  standard deviation (SD) =  $19.28 \pm 1.37$ ) participated in the study (among these, I excluded three people: those who were more than two standard deviations (2SD) in response times (1 participant) and in accuracy rates (2 participants)). All the participants gave informed consent, and were paid for their participation. The study was approved by the Keio University Ethics Committee. The participants were native speakers of Japanese, and the sentences and instructions were given in Japanese. I divided

the participants into two groups at random, one in which diagrams were used (Diagrammatic group) and the other in which diagrams were not used (Linguistic group).

### Stimuli

I presented 82 items in a total six sessions. Of these, 33 items were valid syllogisms, 29 were invalid syllogisms, and 20 were baseline. I used the invalid syllogisms as filler stimuli and included only the valid syllogisms and baselines in the analysis. The syllogisms contained content-free sentences such as “*No A are B*” and were presented in two different logical forms. The two argument forms having a universal negative sentence as one of their premises and having an existential sentence as the other were selected. They have been known as the difficult syllogisms for novices (Dickstein, 1978). In the baseline condition, we used the syllogisms having a trivially true conclusion, i.e., one of the premises can be identified with the conclusion). Examples of each of these categories of stimuli are shown in Fig. 3.13.

### Training

Prior to scanning, the participants received a training session in which the form of a reasoning problem was described as that containing two premises (“*All B are A*” and “*All C are B*”) and a conclusion (“*All C are A*”). The participants were asked to respond “true” if the provided conclusion was valid and “false” if the provided conclusion was not valid. In addition, both the groups were given a few minutes to read one-page instructions on the meaning of the categorical sentences used in the syllogisms (for details, see Appendix B.1.1). Furthermore, the Diagrammatic group was given a few minutes to read two-page instructions on the meaning of diagrams, which did not include any instructions on how to manipulate diagrams in solving the syllogisms (for details, see Appendix B.1.2). The experimenter aimed to resolve any question or misunderstanding without any formal logical instruction.

### Design and procedure

Stimuli were randomly presented in an event-related design. Stimuli presentation in my study is shown in Fig.3.14 . The length of each trial was 32 s (including rest). An asterisk (\*) indicated the start of a trial for 1 s. The sentences appeared

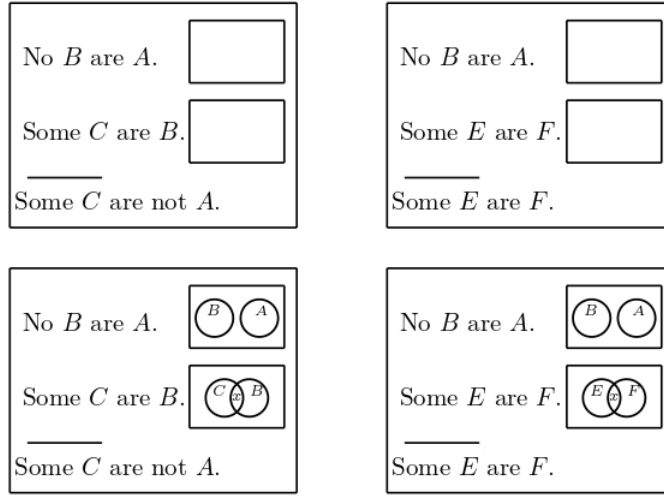


Fig. 3.13 Examples of the four categories of stimuli. A reasoning task of the Linguistic group (upper left side), a baseline condition of the Linguistic group (upper right side), a reasoning task of the Diagrammatic group (lower left side), and a baseline condition of the Diagrammatic group (upper right side).

on the screen one at a time with the first sentence (premise 1) appearing at 1 s, second sentence (premise 2) at 4.5 s, and last sentence (conclusion) at 7.5 s. After the conclusion appeared, both the groups were given 15 s to press the “true” or “false” button for evaluating the sentence. All of the three sentences remained on the screen until the trial has finished (22 s).

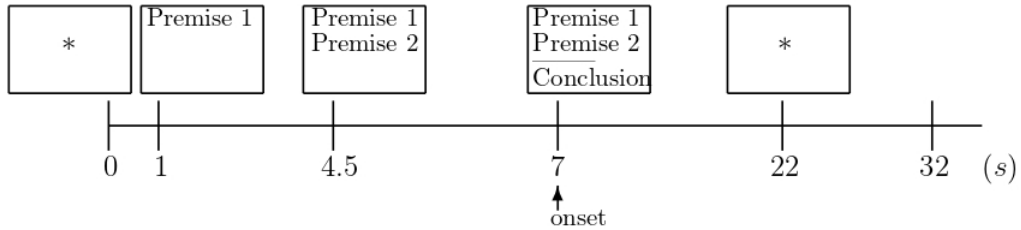


Fig. 3.14 Event-related stimuli presentation

### Data acquisition and analysis

I used a 3T scanner (Siemens Trio Tim; field of view (FOV) = 250 mm, matrix size =  $64 \times 64$ , thickness = 5 mm (gap = 1 mm), Repetition Time (TR)/Echo Time (TE)/Flip angle ( $\alpha$ ) = 3000 ms / 30 ms /  $90^\circ$ ) for data acquisition.



Statistical analyses were conducted on a voxel-by-voxel basis using SPM5 software (The Wellcome Department of Imaging Neuroscience) on a Matlab platform (Mathworks Inc.). Images were converted to the analyze format and subjected to several preprocessing steps, including realignment to the first acquired image, motion correction, and spatial normalization to the Montreal Neurological Institute (MNI) standard anatomical space by means of an echo planar imaging template. The standardized images were smoothed using an 8-mm Gaussian kernel.

Condition effects at each voxel were estimated according to the general linear model, and regionally specific effects were compared using linear contrast. Each contrast produced a statistical parametric map of the t-statistic for each voxel, which was subsequently transformed to a unit normal Z-distribution. The reported activations survived ten voxel-level intensity thresholds of uncorrected  $p < 0.0001$  using a random effect model. Based on the paradigm for fMRI studies of logical reasoning (Goel, Buchel, Frith, & Dolan, 2000; Noveck, Goel & Smith, 2004; for more detail discussion, see Goel, 2007), the blood oxygenation level dependent (BOLD) signal was modeled as a canonical hemodynamic response function (HRF) with a time derivative at the presentation of the third sentence. Duration times in data analysis were set to average persons' reaction times per trial session. The presentation of the two premises and that of motor responses were modeled out of the analysis.

In addition, the MarsBaR SPM toolbox was used to perform statistical analyses of ROI. From SPM clusters in data of each group, functional ROIs in frontal-parietal regions were built as a first step. In a non-overlapping region, each ROI in the two groups was defined as a cluster which exhibits activation of the relevant regions in one group. In an overlapping region, each ROI in the two groups was combined and transformed to the form of intersection between both groups. The mean signal (t-value) obtained in these ROIs was then statistically analyzed (t-test).

### 3.3.2 Results

#### Behavioral data

The average response time of the reasoning tasks was 6603 ms in the Linguistic group and 5017 ms in the Diagrammatic group. The average response time of the baseline tasks was 2904 ms in the Linguistic group and 2454 ms in the

Diagrammatic group. These data were subjected to two-way analysis of variance (ANOVA) for a mixed design. A significant main effect was found for the existence of diagrams [ $F(1, 27) = 8.405, p < 0.01$ ]. In addition, a significant main effect was found for the difference between the inference and baseline tasks [ $F(1, 27) = 99.918, p < 0.001$ ]. Furthermore, there was a significant interaction effect between these two factors [ $F(1, 27) = 3.292$ ] at a reduced threshold of  $p < 0.10$ . Multiple comparison tests were conducted by Ryan's procedure. The results indicated the following: (i) Regarding the reasoning tasks, the response times in the Diagrammatic group were significantly shorter than those in the Linguistic group [ $F(1, 54) = 11.367, p < 0.005$ ]; (ii) regarding the baseline tasks, there was no significant difference in performance between the Diagrammatic and Linguistic groups; and (iii) in both the Linguistic group and Diagrammatic group, the response times of the baseline tasks were significantly shorter than those of the reasoning tasks [ $F(1, 27) = 69.742, p < 0.001$ , for the Linguistic group, and  $F(1, 27) = 33.469, p < 0.001$  for the Diagrammatic group].

The average accuracy rates of the reasoning and baseline tasks were analyzed. In the reasoning tasks, there was no significant difference in performance between the Linguistic and Diagrammatic groups (82.3 % for the Linguistic group and 85.5 % for the Diagrammatic group). Similarly, in the baseline tasks, there was no significant difference in performance between the Diagrammatic and Linguistic groups (more than 98% for both the groups).

### **fMRI data**

In both the Linguistic group and Diagrammatic group, the main effect of reasoning (i.e., minus baseline) activated a large bilateral frontal-parietal network; uncorrected  $p < 0.0001$ . (Tables 3.4–3.5, Fig. 3.15). Only the Linguistic group exhibited activation of the left middle frontal gyrus (near BA 10), whereas the Diagrammatic group did not. On the other hand, both groups exhibited activation of the left inferior PFC (near BA 47), the bilateral dorsal PFC (BA 6), the bilateral parietal lobe (BA 7), and bilateral occipital gyrus. The activations of the left dorsal PFC (mainly BA 6), peaking at  $[-28, -2, 54; -6, 6, 60]$  in the Linguistic group and  $[-40, 12, 36; -26, -2, 40; -50, 30, 32]$  in the Diagrammatic group, and left parietal lobes (BA 7), peaking at  $[-24, -54, 50]$  in the Linguistic group and  $[-28, -60, 42]$  in the Diagrammatic group, and bilateral occipital gyrus were high and survived the intensity thresholds of corrected  $p < 0.05$ .

Table 3.4 Coordinates and Z-scores of the Linguistic group

Location	Brodmann area	MNI Coordinates			Z-score
		X	Y	Z	
Reasoning - baseline					
Lt. middle occipital gyrus	18/19	-26	-92	4	5.50
Lt. inferior occipital gyrus	18/19	-30	-85	-8	5.49
Lt. superior parietal lobe	7	-24	-54	50	5.33
Lt. middle frontal gyrus	6	-28	-2	54	5.01
	6	-30	4	42	4.23
	6	-6	6	60	4.99
	6	-4	18	52	4.38
	10	-32	40	0	4.58
	10	-34	48	0	4.42
	9	-46	28	30	4.35
Lt. superior longitudinal fasciculus	6	-44	10	24	4.18
Rt. middle frontal gyrus	6	32	2	64	4.42
	6	28	2	54	4.28
	6	26	-8	52	4.14
Lt. inferior frontal gyrus	47	-28	22	-5	4.47
Lt. optic radiation		-24	-26	-4	4.55
Lt. lateral globus pallidus		-18	-2	14	4.42
Lt. caudate nucleus		-16	8	14	3.74

Lt., left; Rt., right; MNI, Montreal Neurological Institute

Table 3.5 Coordinates and Z-scores of the Diagrammatic group

	Brodmann	MNI Coordinates			
Location	area	X	Y	Z	Z-score
Reasoning - baseline					
Lt. middle occipital gyrus	18/19	-32	-92	2	5.78
	18/19	-22	-92	14	5.43
Lt. superior parietal lobe	7	-28	-60	42	5.19
Lt. inferior frontal gyrus	6	-40	12	36	5.19
	47	-26	20	-2	4.42
Lt. middle frontal gyrus	9	-50	20	32	5.11
Rt. middle frontal gyrus	6	8	16	50	4.14
Lt. fasciculus occipito-frontalis	6	-26	-2	40	5.06
		-16	10	24	3.99
Lt. sulcus callosomarginalis	6	-6	14	54	4.81
Lt. middle temporal gyrus	20/21	-50	-50	-2	4.71
Lt. inferior temporal gyrus	20/21	-50	-60	-12	4.37
	20/21	-50	-52	-14	4.26
Lt. caudate nucleus		-16	-8	24	4.38

Lt., left; Rt., right; MNI, Montreal Neurological Institute

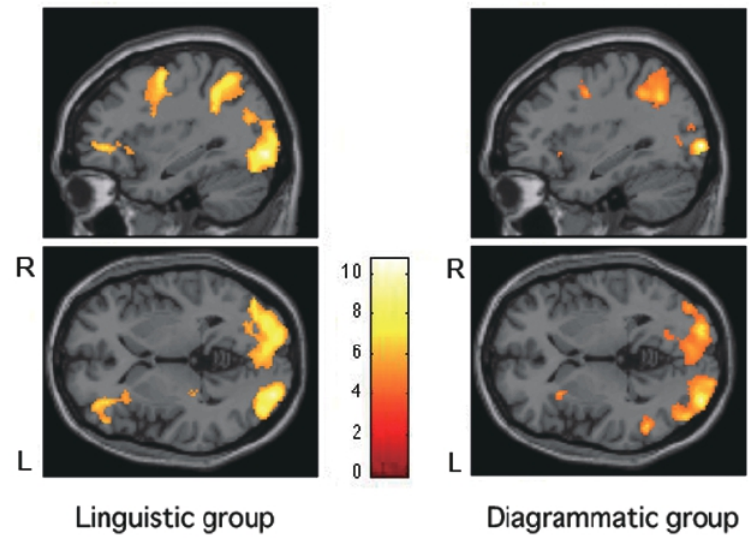


Fig. 3.15 Reasoning minus baseline contrast. Mean group activations in the Linguistic group (left side) and Diagrammatic group (right side); uncorrected  $p < 0.0001$ . R, right; L, left.

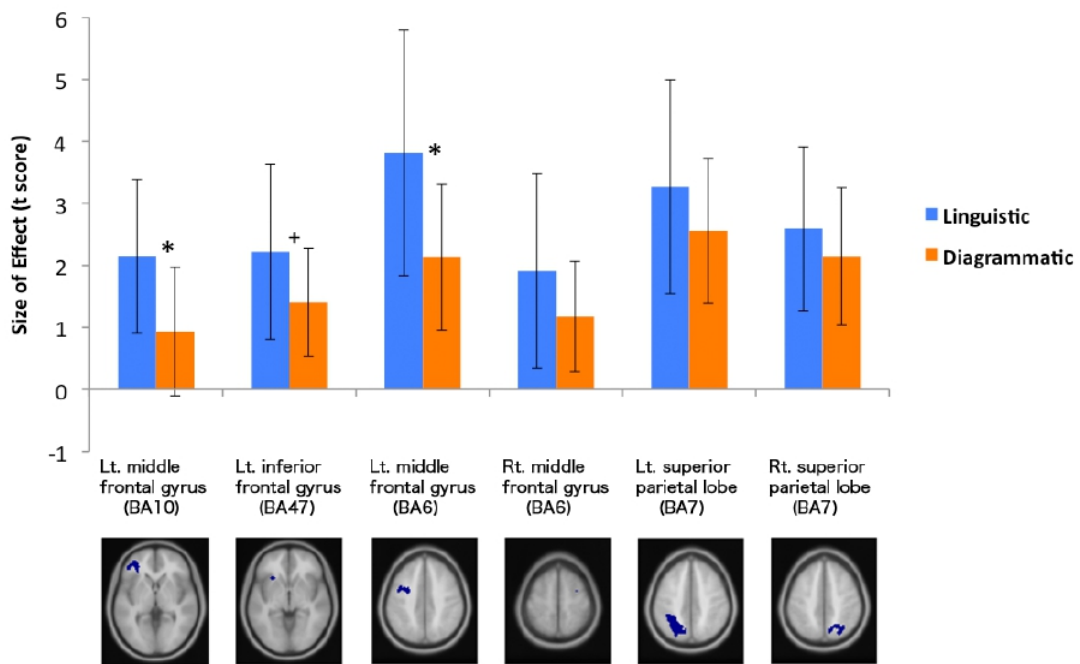


Fig. 3.16 Comparison between the effect sizes of each ROIs in the Linguistic and Diagrammatic groups; Error bars indicate SD. Star refers to the significant difference at the threshold of  $p < 0.05$ . Plus refers to the threshold of  $p < 0.10$ .

To compare the size of activation in the two groups, ROI analysis were conducted, as is shown in Fig 3.16. As a non-overlapping region, analysis of the left inferior frontal gyrus (near BA 10) showed that activation in the Diagrammatic group ( $t = 0.935$ ) was significantly lower than that in the Linguistic group ( $t = 2.153$ ) ( $p < 0.05$ ).

In the evaluation of the overlapping region among the frontal-parietal regions, analysis of the left inferior PFC (near BA 47) showed that activation in the Diagrammatic group ( $t = 1.399$ ) was significantly lower than that in the Linguistic group ( $t = 2.219$ ) at a reduced threshold of  $p < 0.10$ . Analysis of the left dorsal PFC (BA 6) showed that activation in the Diagrammatic group ( $t = 2.132$ ) was significantly lower than that in the Linguistic group ( $t = 3.814$ ) ( $p < 0.05$ ). Similarly, analysis of the right dorsal PFC (BA 6) showed that activation in the Diagrammatic group ( $t = 1.175$ ) was lower than that in the Linguistic group ( $t = 1.907$ ); however, the difference was not statistically significant. In contrast, analysis of the left parietal lobe (BA 7) showed that there was no significant difference between activation in the Diagrammatic Group ( $t = 2.558$ ) and Linguistic group ( $t = 3.269$ ). Similarly, analysis of the right parietal lobe (BA 7) showed that there was no significant difference between activation in the Diagrammatic group ( $t = 2.156$ ) and Linguistic group ( $t = 2.594$ ).

### 3.3.3 Discussion

The present behavioral results, in particular, the response times of the reasoning tasks, indicate that diagrammatic representations are effective and involve computational off-loading during deductive reasoning. In order to investigate whether external diagrams are effective in inferential processes, I analyzed the present functional brain imaging data.

In the present study, complex (linguistic) deductive reasoning revealed activation in the left middle PFC (near rostral PFC: BA 10). As is shown in Fig. 3.16, the cluster of activated voxels extended mainly to the negative direction (-50 mm) of the x-axis, to the positive direction (54 mm) of the y-axis, and to the negative-positive directions (from -14 to 6 mm) of the z-axis from the peaking points [-32, 40, 0; -34, 48, 0]. Activation of the rostral PFC (BA 10) has been observed in the tasks completed by integrating the results of multiple suboperations (peaking at [-32, 50, 9] using Raven's Progressive Matrices (Christoff et al., 2001); peaking at [-36, 60, 8] using Tower of London task (van den Heuvel

et al., 2003); reviewed in Ramnani & Owen, 2004). In a previous study of Badre and D’Esposito (2009), “making a sandwich” has been described as an example of such a task. The action of making a sandwich is the main goal, which can be divided into specific subgoals such as slicing the bread or spreading mayonnaise. In this sense, the tasks such as making a sandwich are abstract.

Activation patterns are known in deductive tasks. For example, activation of the rostral PFC (BA 10), peaking at  $[-44, 42, -5]$ , in syllogistic reasoning tasks has been reported in Rodriguez-Moreno and Hirschb (2009). In addition, activation of the rostral PFC (BA 10), peaking at  $[-46, 50, -4]$ , in relatively complex deductive reasoning tasks (conditional inference with the connectives of negation and disjunction) has been reported in Monti, Parsons, and Osherson (2009). Deductive reasoning is a higher-order cognitive ability that involves combining information gathered from several different sources (see Chapter 1). The present findings indicate that the tasks in standard situations (the Linguistic group) require abstract information processing in the middle PFC (including BA 10). In contrast, in the Diagrammatic group, syllogism solving can replace the tasks of concrete operations of diagrams. In other words, in the Diagrammatic group, the abstract processes constituting suboperations are reduced, and the middle PFC (near BA 10) is therefore not activated.

It should be noted here that not all the imaging studies of deductive reasoning such as categorical syllogism have reported the activation of BA 10 (e.g., Goel, Buchel, Frith, & Dolan, 2000; Goel & Dolan, 2003; Reverberi et al., 2010; 2012). However, as discussed in Prado, Chadha, and Booth (2011), the disagreement about the BA 10 may be explained by the complexities of arguments used in the experiments. We focused on the particular complex syllogisms involving existential sentences to emphasize the efficiency of diagrams in reasoning. On the other hand, Goel, Buchel, Frith, and Dolan (2000), Goel and Dolan (2003), and Reverberi et al (2010; 2012) covered a wide range of syllogistic types including complex and simple ones. Even given the explanation, of course, it is hard to say that the role of BA 10 in deductive reasoning is completely clear, since the general role of BA 10 is highly controversial (for a review, see Ramnani & Owen, 2004; Badre & D’Esposito, 2009). Nevertheless, the present findings of BA 10 could count as a partial evidence for the hypothesis on the inferential efficacy of diagrams. This is considered to provide a consistent view between brain and cognitive models about inferential complexity and efficiency.

Activity in the left inferior PFC such as BA 47 has been involved in semantic processing of language and sentences (cf. Poldrack et al., 1999). Furthermore, neuroimaging studies of syllogistic reasoning have reported brain activation in this region (Rodriguez-Morenoa & Hirschb, 2009; Reverberi et al., 2010; 2012). The findings support the standard two-staged framework in natural language semantics (Blackburn & Bos, 2005) according to which sentences are first associated with semantic information, following which the validity of the argument is checked. In my study, it is possible that compared with the participants in the Linguistic group, the participants in the Diagrammatic group did not depend on semantic processing. In such a view, the syntactical operations such as rule selection and implementation are taken to be meaningful. This holds not only in contentful syllogisms such as “All swans are white”, but also in no-content syllogisms used in my study. In fact, the study of Noveck, Goel and Smith (2004), in which no-content conditional reasoning tasks used, also has reported the activations of the regions including BA 47. Given the findings, my discussion about the role of BA 47 is considered to be consistent with the one in Reverberi et al. (2012), in which the BA 47 is critical for the rule selection and implementation in deductive reasoning.

Many imaging studies of problem solving that involved syllogistic reasoning have shown activation of the ventrolateral PFC (BA 6) (e.g., Goel & Dolan, 2004; Rodriguez-Morenoa & Hirschb, 2009). However, the relationship between BA 6 function and problem solving remains unknown. On the basis of the findings of Monti, Osherson, Martinez and Parsons (2007), which also included brain activation in conditional deductive tasks, activation of BA 6 may be explained by the following two possibilities: (i) increased saccades for complex arguments and (ii) rule-based association as seen in non-motor numeric, spatial, and verbal mental operation tasks (Hanakawa et al., 2002). It seems reasonable to suppose that activation of BA 6 is correlated with manipulation of representations in tasks. On the basis of these findings, decreased activation of BA 6 in the Diagrammatic group may indicate that the participants simplify the form of deductive tasks by replacing diagram manipulations.

The present behavioral and fMRI results show computational off-loading in diagram use. In addition, on the basis of fMRI data of the rostral PFC (BA 10) and the other two regions, abstract deductive tasks replace concrete manipulation of diagrams. These findings support the hypothesis of the efficacy of diagrams in

inferential processes.

The efficacy of diagrams in inferential processes holds true for Euler diagrams used in the present study and for linear diagrams (a linear variant of Euler diagrams, where set relationships are represented by one-dimensional lines, rather than by circles in a plane), but does not necessarily hold true for Venn diagrams (see Experiments 1 and 2). In Euler diagrams, set relationships are expressed by inclusion and exclusion relations between circles. In contrast, Venn diagrams have a fixed configuration of circles and represent set relationships by stipulating that shaded regions denote the empty set. On the basis of the behavioral results, spatial relationships such as inclusion and exclusion relationships can be considered to play an essential role in naturalizing deductive reasoning.

Of note, activation of the parietal lobe (BA 7) has been observed in both the Linguistic and Diagrammatic groups in this study. My conclusion is consistent with previous findings in the literature of cognitive neuroscience. Many imaging studies, in addition to my studies, previous studies have reported parietal lobe (BA 7) activities in deduction and other reasoning tasks and have provided evidence for visuospatial characteristics of linguistic reasoning (e.g., Goel, 2007; Prado, Chadha, & Booth, 2011). Furthermore, these findings were confirmed by Tsujii et al. (2011) using transcranial magnetic stimulation (TMS). They reported that enforced inactivity in parietal lobes impedes the performance of deductive tasks by the participants. In my study, when the participants used diagrams for solving reasoning tasks, parietal lobes were actively involved in visuospatial manipulations and normal linguistic reasoning. As was mentioned before, the common activated regions in different presentation modality could provide a suggestion on the essential process in human reasoning that do not depend on a particular representative way. The fact that the region is parietal lobe (BA 7) is interesting and important when considering the cognitive origin of deduction and logic. It is because my findings are consistent with the findings of graphical reasoning strategies using Euler-like diagrams in Ford (1994), Bacon, Handley and Newstead (2003), and Bucciarelli and Johnson-Laird (1999), in which cognitive strategies of syllogistic reasoning were investigated by analyzing participants' verbal and/or written protocols. Given the psychological-neural findings, it is also interesting that the nature of logic is taken as spatial thinking.



### 3.4 Children’s reasoning with Euler and Venn diagrams: Experiment 4

As stated in the introduction section, over the previous two decades, a number of researchers have shown an interest in the formal and cognitive properties of diagrammatic reasoning (e.g. Glasgow, Narayanan, & Chandrasekaran, 1995). Logic diagrams, such as Euler and Venn diagrams, have been intensively studied using the method of mathematical logic (e.g. Shin, 1994). Based on such logical analyses, the efficacy of diagrammatic reasoning has been explored in the context of cognitive science (e.g. Shimojima, 1996). However, there have been few attempts to apply these logical and empirical findings to the study of the development of children’s reasoning. In particular, little attention has been paid to the cognitive efficacy of diagrams in children’s deductive reasoning.

Based on the findings of Experiment 1, the current section compares children’s performance in syllogistic reasoning tasks externally supported by Euler and Venn diagrams. An earlier version of this work appeared in Sato, Mineshima, and Takemura (2010c). The present study differs from earlier studies where children were provided with substantial training in using diagrams (e.g. Morgan & Carrington, 1944). I hypothesize that Euler diagrams are *self-guiding* in the sense that syntactic manipulation would be expected to be available even to participants without substantial training in the rules or strategies of manipulation, which is not the case for Venn diagrams. This chapter focuses on whether diagrams have efficacy in inferential processes, rather than in sentence interpretation alone (cf. Agnoli, 1991, reporting that logic diagrams can help children avoid misunderstanding of linguistic materials). I predict that even in the case of untrained children’s reasoning, performance would be better when using Euler diagrams than when using Venn diagrams. Such a result would provide evidence for syntactic manipulation of diagrams in children’s reasoning, giving a partial explanation for the efficacy of diagram use in children’s deductive reasoning.

#### 3.4.1 Method

##### Participants

Eighty-six children in a Japanese public elementary school took part in the experiment. (i) 29 children were in the fourth grade (9- to 10-years-olds) and were divided into the Euler group (15 children) and the Venn group (14 children). (ii)

31 children were in the fifth grade (10- to 11-years-olds) and were divided into the Euler group (15 children) and the Venn group (16 children). (iii) 26 children were in the sixth grade (11-years-olds) and were divided into the Euler group (13 children) and the Venn group (13 children).

## Materials and Procedure

The experiment was conducted in booklet form. Children were first provided with instructions about the meaning of diagrams (details of the instructions are given in Appendix B.1.5 and B.1.6), then required to solve syllogistic reasoning tasks with diagrams. The instruction and tasks were presented in Japanese. The following explanation of reasoning tasks was presented to participants: *Now we are casting roles in a play. Because we do not have enough people, one person need cast several roles.* The premises and conclusions of the syllogisms were universally quantified sentences, which took the form of either *All A are B* or *No A are B*, where concrete terms appeared in *A* and *B*. Before the syllogistic reasoning task, a pretest was conducted to check whether the children understood the instructions correctly. The children in the Euler group and Venn group were presented with four diagrams listed in Fig. 3.17 and Fig. 3.18, respectively.

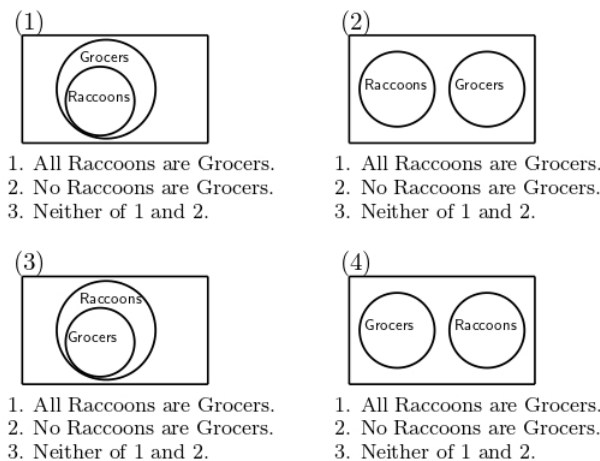


Fig. 3.17 Euler diagrams used in the pretest for children

The highest possible score on the pretest was four, and the cutoff point was three. The Euler group was required to solve syllogisms with Euler diagrams (see Fig. 3.19). The Venn group was required to solve syllogisms with Venn diagrams (see Fig. 3.20). The children were presented with two premises and asked to select

the correct answer from a list of three possibilities: 1. *All A are B*, 2. *No A are B*, and 3. *Neither of 1 and 2*. Both groups were presented with 12 syllogisms, out of which five syllogisms had a valid conclusion (AA1, AE2, AE4, EA1, and EA2 types) and seven had no valid conclusion (AA2, AA3, AA4, AE1, AE3, EA3, and EA4 types). Each task was presented in a random order.

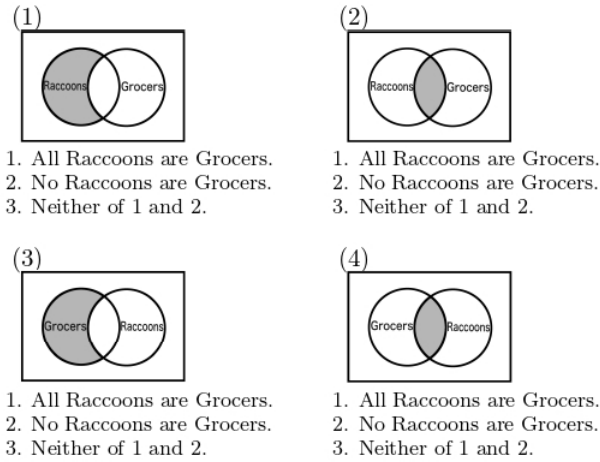


Fig.3.18 Venn diagrams used in the pretest for children

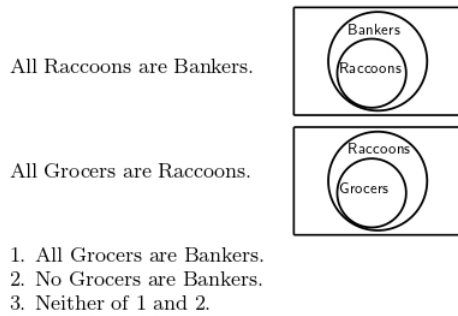


Fig.3.19 An example of a reasoning task of the Euler group (correct answer: 1).

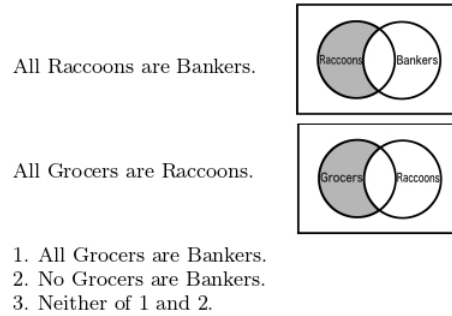


Fig.3.20 An example of a reasoning task of the Venn group (correct answer: 1).

### 3.4.2 Results

#### Pretest

At the fourth-grade level, twelve in the Euler group and one in the Venn group passed the pretest. At the fifth-grade level, eleven in the Euler group and four

in the Venn group passed the pretest. At the sixth-grade level, ten in the Euler group and five in the Venn group passed the pretest. These results indicate that for all grade levels, Venn diagrams are more difficult to understand than Euler diagrams. Since so few of the fourth-grade children could understand the meaning of Venn diagrams correctly, the following statistical analysis of the Venn group begins at the fifth-grade.

### Syllogistic reasoning tasks

The results are given in Table 3.6, where **valid** means syllogism with valid conclusions and **invalid** means syllogisms having no valid conclusion. At the sixth-grade level, the difference in the performance of valid syllogisms between the two groups was significant: 70.0% for the Euler group and 36.0% for the Venn group ( $F(1, 14) = 4.950$ ,  $p < .05$ ). At the fifth-grade level, the performance of valid syllogisms in the Euler group was better than that of the Venn group, although the difference was not statistically significant. At all grade levels, the performances of invalid syllogisms in both groups were lower than the chance level.

Table 3.6 The average accuracy rates of syllogistic reasoning tasks (the bold-types refer to the significant difference between the sixth-grade groups at the level of  $p < .05$ .)

	THE FOURTH GRADE		THE FIFTH GRADE		THE SIXTH GRADE	
	valid	invalid	valid	invalid	valid	invalid
Euler group	48.3%	11.9%	56.3%	29.8%	<b>70.0%</b>	17.1%
Venn group	—	—	35.0%	26.4%	<b>36.0%</b>	22.8%

### 3.4.3 Discussion

The present results showed that the difference between Euler and Venn diagrams had significant effect on the performance of the children in the six grade with respect to the valid syllogisms. This could be interpreted to suggest that the syntactic manipulations of Euler diagrams could be naturally triggered for them, facilitating the processes of combining information in premises. The overall results present evidence for the claim that Euler diagrams do help children solve syllogisms. By contrast, in the case of Venn diagrams, even the performance

of the children who passed the pretest was not so higher than the chance level. This could be explained by supposing that syntactic manipulations of diagrams were not available to the children in the Venn group, and hence they had to rely on usual processes of linguistic inferences, resulting in poor performance in syllogisms.

Regarding the invalid syllogisms, the results indicate that the diagrams did not improve the children's performance. This shows a striking contrast to adults' performance reported in my previous study, where the performance in the invalid syllogisms was significantly improved when participants were presented with Euler diagrams (see Experiment 1). This might be related to the fact that invalid syllogisms usually involve some kind of indeterminacy with respect to the information contained in premises. In the case of reasoning with Euler diagrams, unless some special convention is introduced, such an indeterminacy usually requires enumerating the possible configurations of circles and thus multiplying the conclusion diagrams. In neuroimaging studies, Goel et al. (2007) reported that the predominant activation in the right prefrontal cortex (PFC) was observed in the case of reasoning with indeterminate forms. Shaw et al. (2006) reported that the PFC matures late in development. Children might have an inevitable disadvantage for judging invalidity even when using diagrams. These issues and other are left for future work.

## Chapter 4

# Theoretical–empirical study of the interpretational efficacy

This chapter focuses on a theoretical and empirical analysis of interpretations of sentences and diagrams. The conditions in which diagrams can have interpretational efficacy are further explored. More specifically, the process of information extraction from Euler and Venn diagrams is discussed through experiments (introduced in Sato, Mineshima, & Takemura, 2011), examining the performance of matching times between categorical sentences and corresponding diagrams.

### 4.1 Extracting information from Euler and Venn diagrams

In traditional studies of diagrams in the field of logic and cognitive science, particular emphasis has been placed on the comparison between diagrammatic and sentential (linguistic) representation systems. In addition to the availability of concrete manipulations, a number of properties have been proposed to distinguish diagrammatic from sentential representations, seeking to account for the general advantages of diagrammatic representations have over sentential representations (e.g. Shimojima, 1999b; Stenning, 2000). In contrast, relatively little attention has been paid to the comparison between different diagrammatic representation systems. However, such a comparison might be important for providing a more fine-grained analysis of efficacy of various diagrams in human problem solving.

As a crucial example, consider the process of solving the syllogism we saw above with Venn diagrams, as shown on the right in Fig. 4.1. In Venn diagrams, every circle partially overlaps each other, and meaningful relations among

circles are expressed by *shading* under the convention that shaded regions denote the empty set. Given this semantic information, the process of composing two premise diagrams automatically yields the information corresponding to the correct conclusion, in a similar process to that of Euler diagrams. Intuitively, however, reasoning with Venn diagrams appears to be relatively more difficult. As emphasized in Gurr, Lee and Stenning (1998), what differentiates the two cases is the process of *interpreting* (externally given or internally constructed) diagrammatic representations. In particular, a conventional device such as shading might cause complication in the interpretative processes of Venn diagrams. This suggests that to obtain a more comprehensive account of diagrammatic reasoning, we need to take a closer look at the cognitive processes underlying *information extraction* from diagrams.

The notion of “similarity” has played a traditional role in accounting for differences in the efficacy of semantic aspects of sentential and diagrammatic representations. As mentioned in Section 3.1 of Chapter 2, if a certain *structural correspondence* holds between a representation and what it represents, the representation should be effective in interpretation and communication even for users who have not explicitly learned conventions governing its use. In the literature, the notion of structural correspondence has been specified in various researches (e.g., Barwise & Etchemendy, 1991; Barwise & Hammer, 1996; Gurr, Lee, & Stenning, 1998; Gurr, 1998; 1999; Gattis, 2004). However, a precise characterization of the notion of structural correspondence that could be applied to varieties of diagrammatic representations remains to be explored. In particular, the way in which a cognitive account of interpretation processes could be connected to the formal semantics of diagrams remains unclear.

The question of information extraction has also been investigated in the study of the cognitive role of relatively simple diagrams, such as charts, maps, and graphs (e.g., Ratwani, Trafton, & Boehm-Davis, 2008; Shimojima & Katagiri, 2010). What plays an important role here is the distinction between lower-level and higher-level information. For example, a scatter plot contains lower-level information about specific data and the overall distribution of dots delivers the higher-level information about the structural properties of data (Kosslyn, 1994). However, application of these findings to an analysis of higher cognitive processes, such as deductive reasoning, has not been fully explored.

The present chapter aims to investigate the cognitive processes of interpreting

diagrammatic representations underlying deductive reasoning, combining the insights from these different research traditions. In particular, based on a semantic analysis of diagrams, I argue that a certain structural correspondence between a diagrammatic representation and its semantic contents plays a crucial role in both interpretation and inference processes with the representations. The present approach can also provide a further *constraint* on the choice of different ways of formalizing the abstract syntax and semantics of diagrammatic representations, providing motivation for developing a more appropriate way of approaching the logical study of diagrams. This could contribute to the development of a closer connection between the logical and cognitive approaches to human problem solving with diagrammatic representations.

## 4.2 Hypothesis: The structural correspondence of interpretations

A major goal of the present chapter is to explore the hypothesis that the “matching” relation between the diagrammatic representation used in deductive reasoning and the conveyed information available to users plays an important role in effective diagrammatic reasoning. As a case study, we focus on the use of logic diagrams in syllogistic reasoning. To clarify the relevant structural relationship between logic diagrams and their semantic contents, it is helpful to first identify what semantic information is carried by syllogistic sentences, using the insight obtained via the semantics of natural language (see also Chapter 2).

**1. The relational analysis of quantified sentence.** As discussed in Section 2 of Chapter 2, syllogistic reasoning, as investigated in the psychology literature, can be regarded as a special case of reasoning with *quantificational* sentences in natural language. According to the conventional view, such sentences are analyzed using representations in first-order predicate logic, which essentially involve quantification over *individuals* as semantic primitives. In the field of natural language semantics, in contrast, quantifiers in natural language, such as *all*, *some* and *no*, are analyzed as denoting *relations* between sets (i.e., *generalized quantifiers*; Barwise & Cooper, 1981). Thus, a sentence of the form *All A are B* is analyzed as  $\mathbf{A} \subset \mathbf{B}$ , rather than as the first-order representation  $\forall x(Ax \rightarrow Bx)$ . Similarly, *No A are B* can be analyzed as expressing  $\mathbf{A} \cap \mathbf{B} = \emptyset$ . Here, the semantic primitives of quantificational sentences are considered as the relations



between sets, such as subset relation and disjointness relation. It should be emphasized that such a “relational” formulation of the meaning of a quantified sentence could capture not only the truth condition or logical form of a sentence, as is traditionally assumed, but could also capture *how* readers mentally represent such a truth condition or logical form (e.g., Politzer, van den Henst, Luche, & Noveck, 2006). The relational approach to quantificational sentences has also been successfully applied to the psychological study of deduction, resulting in a processing model based on the assumption that reasoning with quantifiers are performed in terms of sets rather than individual items (Geurts, 2003a). For a discussion on the general role of relational knowledge in higher cognition, see Halford, Wilson, & Phillips (1998, 2010).

In summary, it is reasonable to assume that the semantic primitives of quantificational sentences in natural language are *relations* between sets, and that people’s reasoning with quantified constructions are sensitive to such a relational structure.

**2. Relational analysis of Euler diagrams.** In the logical study of Euler and Venn diagrams (e.g., Shin, 1994), diagrammatic representations have been given their own formal syntax and semantics, in a similar way to formulas in mathematical logic. What is remarkable here is that a diagram may have several “equivalent” formalizations as illustrated in Section 3 of Chapter 2. As an illustrative example, reconsider a simple Euler diagram  $E$  in Fig. 4.1. This diagram

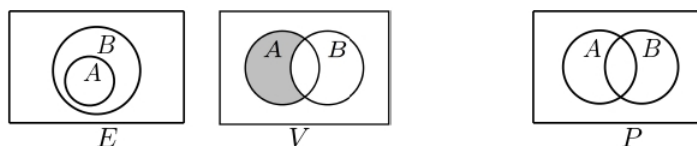


Fig. 4.1 Examples of an Euler diagram  $E$  and a Venn diagram  $V$  that correspond to the sentence *All A are B*. The diagrams  $P$  is a so-called “plain” diagram, which expresses a tautology.

can be intuitively interpreted as denoting the subset relation between sets  $\mathbf{A}$  and  $\mathbf{B}$ , i.e.,  $\mathbf{A} \subset \mathbf{B}$ . However, it can also be interpreted as expressing that  $\mathbf{A} \cap \overline{\mathbf{B}} = \emptyset$ , where  $\overline{\mathbf{B}}$  denotes the complement of the set  $\mathbf{B}$ . Correspondingly, there are two ways of formalizing the abstract syntax of Euler diagrams (Mineshima, Okada & Takemura, 2010; 2012a). According to the “relational” approach, Euler diagrams are abstractly specified as a set of topological relations holding between

objects in the diagrams. For example, the diagram  $E$  in Fig. 4.1 is represented as  $\{A \sqsubset B\}$ , where  $A \sqsubset B$  means that circle  $A$  is inside circle  $B$ . Another approach is a “region-based” approach, which is fairly standard in the logical study of diagrammatic reasoning (e.g. Howse, Stapleton, & Taylor, 2005). Here, diagrams are abstractly defined in terms of regions and emptiness. Thus the diagram  $E$  in Fig. 4.1 can be represented by specifying the region  $(A, \overline{B})$ , the region inside circle  $A$  and outside circle  $B$ , as a “missing” region.

Clearly, these two ways of defining Euler diagrams can be used to predict, for each diagram, the equivalent truth-condition. The difference arises from the way these truth-conditions are given, raising the question of which of these (or other) formulations reflects the way the user represents the semantic content of a given diagram? Here the region-based formulation appears to be more natural for the meaning of the Venn diagram, such as diagram  $V$  shown in Fig. 4.1. Given the convention that the shaded region denotes the empty set,  $\mathbf{A} \cap \overline{\mathbf{B}} = \emptyset$  has the syntactic reading “the region inside the circle  $A$  but outside the circle  $B$  denotes the empty set”, or, less formally, “there is nothing which is  $A$  but not  $B$ ”.<sup>1</sup> The two ways of formulating logic diagrams can thus be summarized as follows.

Table 4.1 The two ways of formulating logic diagrams

	building blocks	meaningful units (semantic primitives)
Relation-based analysis	circles	relations between circles
Region-based analysis	regions	non-emptiness of minimal regions

Here, the basic hypothesis is that an Euler diagram like  $E$  in Fig. 4.1 expresses relational information that could be accounted for by relation-based analysis, triggering a relational representation such as  $\mathbf{A} \subset \mathbf{B}$  to the users. In contrast, Venn diagrams are subject to region-based analysis, triggering semantic information

<sup>1</sup>It should be noted here that throughout this discussion, it is assumed that both Euler and Venn diagrams adopt the convention that each *unshaded* region lacks existential imports, i.e., that the region may denote an empty set. Thus, the diagram  $P$  in Fig. 4.1, where the circles  $A$  and  $B$  partially overlap each other, conveys semantically tautologous information. In other words, this diagram implies that the semantic relationship between  $A$  and  $B$  is indeterminate. In this respect, the current conception of semantics differs from that discussed by Johnson-Laird (1983; for a more detailed discussion of the semantics of Euler and Venn diagrams, see Hammer & Shin, 1998 and Section 2.1 of Chapter 2 of the present thesis).

such as  $\mathbf{A} \cap \overline{\mathbf{B}} = \emptyset$  to their users.

Above, it was shown that syllogistic sentences are quantificational sentences of the relational form, schematically represented as  $Q(A, B)$ , and that such sentences force a reasoner to construct, and operate on, relational representations in reasoning. If the basic hypothesis presented above is correct, Euler diagrams directly express the topological relationship between circles. Thus, it is hypothesized that when a reasoner is asked to match a syllogistic sentence with a corresponding Euler diagram (or vice versa), they could appeal to the process of immediately reading off the relational information from an Euler diagram (i.e., without any intermediate steps), then verify that it is the same as the information conveyed by the sentence. If the analysis is correct, the structural correspondences between representational elements may hold in the categorical sentence in Euler diagrams. A model of the matching processes between sentence and diagram is shown in Fig. 4.2, in which the processes of Euler diagrams is described on the left side.

In contrast, Venn diagrams have a fixed configuration of circles and represent set relationships *indirectly* by stipulating that shaded regions denote an empty set. Accordingly, the process of extracting the relevant relational information from a Venn diagram would be expected to proceed in several steps. As a concrete example, consider how a reasoner could extract the correct relational representation from the Venn diagram in Fig. 4.2 (right side) above. If a region inside of some circles and outside of the rest of the circles (possibly none) in a diagram is referred to as a *minimal* region, it can be seen that the diagram has eight minimal regions. First, the reasoner needs to check each minimal region to determine whether it is shaded or not, as lower-level information. In this example, the regions  $(A, \overline{B}, \overline{C})$ ,  $(A, \overline{B}, C)$ ,  $(A, B, \overline{C})$ , and  $(\overline{A}, B, C)$  are shaded. Next, the reasoner internally builds the *segments* that combine the shaded minimal regions continuous with each other, as well as those which combine the unshaded minimal regions. This step makes it possible for the reasoner to conclude that the diagram delivers higher-level information “There is nothing which is  $A$  but not  $C$ ”. The reasoner could then paraphrase this as “All  $A$  are  $C$ ”, corresponding to the required information in syllogistic inferences. Thus, it is hypothesized that such complexities would cause some difficulties in extracting the required information from Venn diagrams.

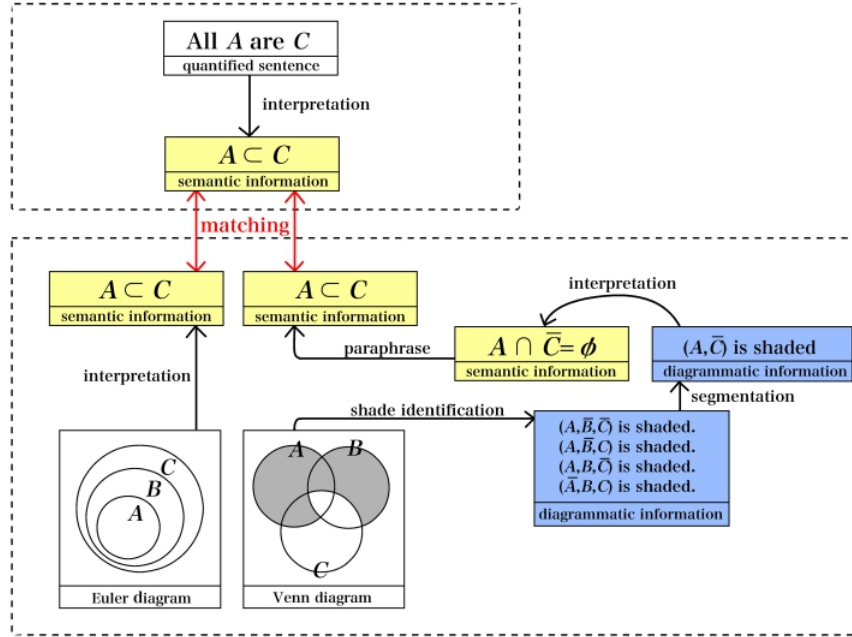


Fig. 4.2 Cognitive model for matching processes between sentence and diagram

## 4.3 Sentence-diagram matching test: Experiment 5-1

As an initial test of my hypothesis, I conducted a “sentence-diagram” matching test, in which participants were presented with a syllogistic sentence and asked to choose the diagram expressing the same information. In Experiment 5-1, we used a simple form of Venn diagrams consisting of two circles (see the diagram *V* in Fig. 4.1), rather than three. In order to exclude external factors such as familiarity with presented diagrams, participants were provided with sufficient informal explanation of the semantics of diagrammatic representations.

### 4.3.1 Method

#### Participants

Twenty-seven undergraduates and graduates (mean age  $22.34 \pm 3.27$  SD) took part in the experiment. Participants gave written informed consent before taking

part in the experiment, and were financially reimbursed for participating. All participants were native speakers of Japanese, and task sentences and instructions were presented in Japanese. The participants were divided into two groups: the Euler group ( $N = 13$ ) and the Venn group ( $N = 14$ ).

## Materials

The syllogistic sentences used in the experiment can be divided into *existential* and *non-existential* sentences, with the following patterns:

Table 4.2 Categorical sentences used in the matching test

Non-existential sentences	Existential sentences
(1) All $A$ are $B$ .	(5) Some $A$ are $B$ .
(2) All $B$ are $A$ .	(6) Some $B$ are $A$ .
(3) No $A$ are $B$ .	(7) Some $A$ are not $B$ .
(4) No $B$ are $A$ .	(8) Some $B$ are not $A$ .

The participants were presented with one sentence on a computer monitor, and were instructed to choose the corresponding diagram (if any). Fig. 4.3 shows the templates of the tasks for the two groups. Here, in both diagrams, the sentence *All  $A$  are  $B$*  corresponds to Answer 1, *No  $A$  are  $B$*  and *No  $B$  are  $A$*  to Answer 2, *Some  $A$  are  $B$*  and *Some  $B$  are  $A$*  to Answer 3, and *Some  $A$  are not  $B$*  to Answer 4.<sup>2</sup> Stimuli were presented in a random order. After a task sentence and four diagrams appeared, participants were instructed to press one of five buttons to indicate their response. There was no time limit for solving the matching tests.

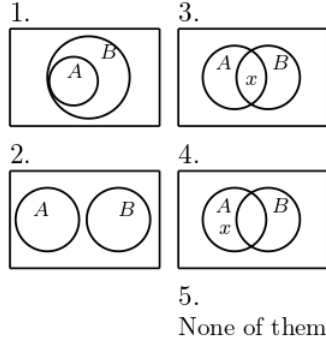
## Procedure

The experiment was conducted individually in a quiet room.

(1) **Instruction and pretest.** Before the test, participants were provided with instructions about the meaning of the sentences and diagrams used (see Appendixes B.1.1; B.1.2; B.1.3). A pretest was conducted to check whether participants understood the instructions correctly; participants were presented with ten

<sup>2</sup>It should be noted that a diagram in which circles partially overlap each other does not express any specific semantic relationship between them (see the diagram  $P$  in Fig. 4.1). To express the existence of objects (i.e., the non-emptiness of a set), I use the point  $x$ . As a consequence, in Euler and Venn diagrams, existential sentences are represented in the same way as indicated in Fig. 4.3. Note also that the sentences *All  $B$  are  $A$*  and *Some  $B$  are not  $A$*  have no corresponding diagram, hence the correct answer is 5 (“None of them”).

*A syllogistic sentence  
is inserted in this area.*



*A syllogistic sentence  
is inserted in this area.*

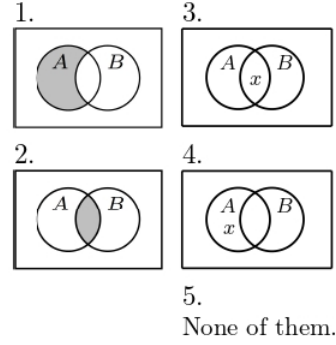


Fig. 4.3 Templates of task sentences with Euler diagrams (left) and Venn diagrams (right) used in the experiment.

diagrams of basic forms and instructed to choose all the sentences (if any) that conveyed the same meaning as the diagrams. After the pretest, the experimenter gave the participants feedback indicating whether they answered correctly. When an incorrect answer was given, participants were asked to reread the instructions and select the correct answer.

(2) **Matching task.** A task example was displayed on the monitor. A total of eight different sentences were prepared. The participants were asked to press the button corresponding to their response, as quickly and as accurately as possible.

### 4.3.2 Prediction

For non-existential sentences, it was predicted that the response time to choose Euler diagrams would be shorter than the response time to choose Venn diagrams. For existential sentences, Euler and Venn diagrams took the same form (see Fig. 4.3 above). Hence, it was predicted that there would be no difference between the two cases.

### 4.3.3 Results

Of the 27 participants, I excluded four (one in the Euler group and three in the Venn group), who gave one or less correct answers.

Table 4.3 shows the average response times in the sentence-diagram matching

task for the Euler and Venn diagram groups.

Table 4.3 Response times of the sentence-diagram matching task with Euler and Venn diagrams.

	<i>non-existential sentence</i>	<i>existential sentence</i>
Euler diagrams	07.286s	09.298s
Venn diagrams	11.057s	10.127s

These data were analyzed using a two-way ANOVA with a mixed design. There was no significant main effect of the difference between Euler and Venn diagrams ( $F(1, 21) = 2.203, p > .10$ ). There was no main effect of the difference between existential and non-existential sentences ( $F(1, 21) = 0.484, p > .10$ ). There was a significant interaction effect between these two factors ( $F(1, 21) = 3.575, p < .10$ ). Multiple comparison tests were conducted using Ryan's procedure. The results indicated that (i) regarding non-existential sentences, response times in the sentence-diagram matching task were significantly shorter for Euler diagrams than for Venn diagrams ( $F(1, 42) = 4.730, p < .05$ ), and that (ii) regarding existential sentences, there was no significant difference in performance between the Euler group and the Venn group ( $F(1, 42) = 0.270, p > .10$ ). It should be noted that the average accuracy rates for both types of diagrams were very high (more than 82 %). Furthermore, no significant difference resulted from changing the order of the terms in the presented sentences (e.g., between *No A are B* and *No B are A*).

Overall, the results provide partial evidence for the hypothesis that the process of extracting relational information from Euler diagrams to match it with sentence meaning would be simple and immediate, whereas in the case of Venn diagrams it appeared to be more complicated.

## 4.4 Diagram-sentence matching test: Experiment 5-2

To further test my hypothesis, I conducted a "diagram-sentence" matching test, in which participants were presented with a diagram and asked to select the sentence conveying the same information. In Experiment 5-2, I used Euler and

Venn diagrams consisting of *three* circles as in Fig. 4.2. This method was expected to be more sensitive to the difference in complexity of information-extracting processes for the two types of diagrams.

#### 4.4.1 Method

##### Participants

Twenty-three undergraduate and graduate students (mean age  $22.73 \pm 2.41$  SD) took part in the experiments. Participants gave written informed consent before taking part in the experiments, and were financially reimbursed for their participation. The participants were native speakers of Japanese, and the task sentences and instructions were presented in Japanese. The participants were divided into two groups: Euler group ( $N = 12$ ) and Venn groups ( $N = 11$ ).

##### Materials

Eleven Euler diagrams and 11 corresponding Venn diagrams were used in the “diagram-sentence” matching task, as shown in Fig. 4.4 and Fig. 4.5. Task examples for Euler and Venn groups are shown in Fig. 4.6. Participants were presented with 11 Euler diagrams (Venn diagrams), and instructed to choose all sentences (if any) that expressed the same information as a given diagram. There were five options: *All-*, *No-*, *Some-*, *Some-not*, and *None of them*, as indicated in Fig. 4.6. Stimuli were presented in a random order. When the task diagrams and sentences appeared, participants were instructed to press the button corresponding to their response, as quickly and as accurately as possible. There was no time limit for solving the task.

##### Procedure

The experiment was conducted in the same manner as Experiment 5-1. Instructions about sentences and diagrams were provided, pretests were conducted, and the matching task was performed.

#### 4.4.2 Prediction

It was predicted that when diagrams did not contain a point  $x$ , the response time for Euler diagrams would be shorter than that for Venn diagrams. When Venn diagrams contained a point  $x$ , users needed to recognize the relationship between



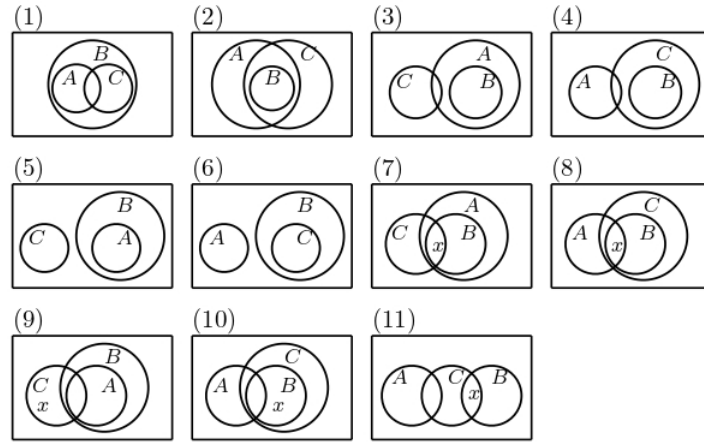


Fig. 4.4 The Euler diagrams used in Experiment 5-2

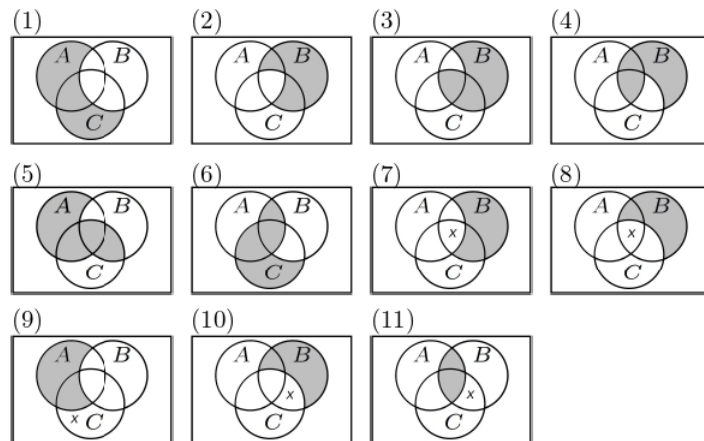


Fig. 4.5 The Venn diagrams used in Experiment 5-2

the point  $x$  and a relevant circle. In such cases, the relationship between circles is irrelevant, so the processes of identifying each minimal region as shaded or unshaded and constructing the relevant segments could be omitted. As such, the response time for Venn diagrams containing a point  $x$  would be expected to be shorter than that for diagrams that did not. In contrast, when an Euler diagram contained a point  $x$ , there was no difference with respect to whether it contained a point  $x$  or not, since in both cases the users needed to check the relationship between two objects, i.e., the relationship between two circles or the relationship between a circle and a point.

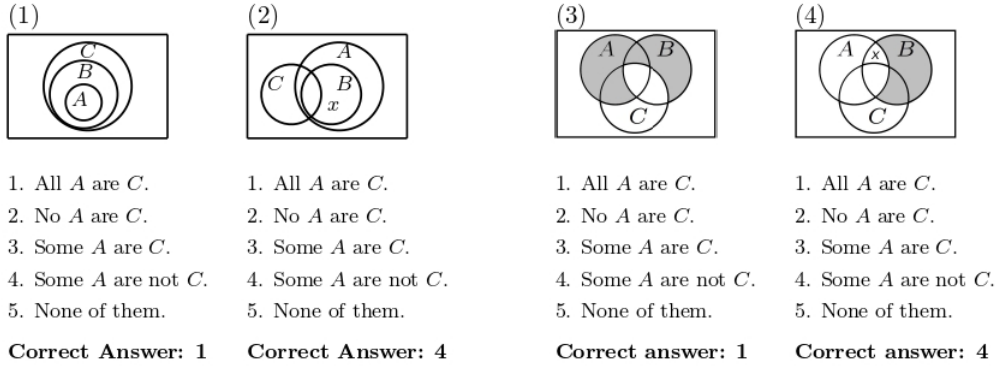


Fig. 4.6 Examples of the diagram-sentence matching task with Euler diagrams (1) (2) and Venn diagrams (3) (4).

### 4.4.3 Results

Table 4.4 shows the average response times in the “diagram-sentence” matching task with Euler and Venn diagrams. In the table, “no-point” refers to diagrams that did not contain point  $x$  and “point” stands for those that contained point  $x$ . These data were analyzed with a two-way ANOVA with a mixed design.

Table 4.4 Response times for the “diagram-sentence” matching tasks with Euler and Venn diagrams.

	<i>no-point</i>	<i>point</i>
Euler diagrams	10.137s	11.946s
Venn diagrams	20.435s	14.022s

There was a significant main effect for the difference between Euler and Venn diagrams, ( $F(1, 21) = 6.087, p < .05$ ). There was no significant main effect for the difference between diagrams containing points and those containing no points, ( $F(1, 21) = 2.032, p > .10$ ). There was a significant interaction effect between these two factors, ( $F(1, 21) = 6.480, p < .05$ ). Multiple comparison tests were conducted using Ryan’s procedure. The results indicated that (i) regarding diagrams without point  $x$ , the response times for Euler diagrams were significantly shorter than those for Venn diagrams ( $F(1, 42) = 11.919, p < .005$ ), and that (ii) regarding Venn diagrams, the response times for diagrams containing a point  $x$  were significantly shorter than those not containing a point ( $F(1, 21) = 7.885,$

$p < .05$ ). (iii) regarding Euler diagrams, there was no significant difference in response times between cases that contained a point  $x$  and those that did not ( $F(1, 21) = 0.627, p > .10$ ). These results clearly support my predictions. It should be noted that the average accuracy rates for both diagrams were very high (more than 83 %).

## 4.5 Discussion

Overall, the results of the two experiments in this chapter provide evidence for the proposed multiple stage view of information-extraction from Euler and Venn diagrams. According to this view, a process of extracting relational information from Euler diagrams consists of a single step, whereas that from Venn diagrams consists of multiple steps, i.e., identifying shaded and unshaded minimal regions and constructing the segments corresponding to the terms in question. These claims are consistent with the empirical findings of Fish, Khazaei, and Roast (2011), where the difficulties in comprehending the convention of shading in Venn diagrams were reported. Regarding the condition of interpretational efficacy of diagrams, the result could count as support for the hypothesis that the structural correspondences between representational elements hold for categorical sentences and Euler diagrams.

Together with these previous findings, the present results suggest the possibility of a model of the overall processes of reasoning with diagrams for deduction, where not only the availability of syntactic manipulations of diagrams but also a subtle difference in the process of extracting information from diagrams might influence the effectiveness of diagrammatic inference. Importantly, the present cognitive study may provide helpful additional constraints or data for theorizing about proof theory (syntax) and model theory (semantics) for higher-level diagrammatic representations. Cognitive or procedural differences in semantically equivalent ways of specifying truth-conditions or logical forms of diagrams (and sentences) tend to be neglected in theorizing about formal semantics and proof theory for representations. Even in the study of diagrammatic logic, such theoretical investigations could be conducted independently of any cognitive considerations. The present results suggest the possibility of a more constrained and integrated framework to deal with both logical and cognitive aspects of diagram use.

# Chapter 5

## Concluding discussion

In Section 1, I summarize the findings of my experiments, then reconsider the efficacy of Euler diagrams by examining the way in which spatial simulations make relational information available. Here we find answers to the questions posed in Section 3 of Chapter 1. Section 2 considers the correspondence relation between representation systems (what is called *structural* correspondence) and the correspondence relation between inference systems (what is called *procedural* correspondence) as conditions for realizing the two types of efficacy: interpretational and inferential efficacy. Section 3 compares the theoretical and experimental findings in this thesis with related studies in cognitive science. Specifically, the graphical method of case identification and the dual process theory of reasoning are discussed. Finally, in Section 4, ideas for future research are discussed.

### 5.1 Conclusion: The efficacy of diagrams in logical reasoning

We found that syllogistic reasoning performance was significantly better with Euler diagrams than with Venn diagrams (Chapter 3, Section 1). Importantly, the same pattern of results was found using linear diagrams that are one-dimensional variants of Euler diagrams (Chapter 3, Section 2). Here, Euler diagrams and Venn diagrams are equivalent in terms of semantic information, and could have *interpretational efficacy* in the current experimental design. Thus, the experimental results provided evidence for the hypothesis that Euler diagrams have *inferential efficacy*, suggesting that the syntactic manipulation of diagrams can be naturally triggered in participants without prior knowledge of inference rules or strategies

(this is the answer to the theoretical question (i)). Two additional findings for both Euler diagrams and the linear variant also support this claim. First, syllogism solving performance with Euler diagrams was significantly better when the convention of crossing (which requires more effort in reasoning processes) was involved, compared with when it was not. Second, even in NVC tasks, syllogistic reasoning performance with Euler diagrams was significantly better than with Venn diagrams. As examined in Section 3 of Chapter 3, changes in activation of the frontal gyrus, including the rostral prefrontal cortex (BA 10), which is implicated in manipulating abstract representations, were observed when comparing hemodynamic changes during reasoning in cases in which Euler diagrams were used with cases in which diagrams were not used. This finding lends further support to the hypothesis of inferential efficacy.

As discussed in Chapter 2, these findings could be explained by *relational structures*, such as inclusion and exclusion, being properties common to categorical syllogistic and Euler diagrammatic inference, in the following way: (1) categorical syllogistic reasoning is concerned with the (inclusion and exclusion) relations between sets. (2) Euler diagrams aid understanding of the underlying semantic relations implicit in categorical sentences in terms of their spatial/geometrical properties (e.g., inclusion and exclusion relations), as examined in Chapter 4. Such diagrams also prevent the misunderstanding of categorical sentences, and hence lead to interpretational efficacy. (3) Moreover, Euler diagram manipulation aids the understanding of relational inferences that are implicit in a categorical syllogism; the spatial relations of Euler diagrams are governed by natural constraints (such as C1 and C2), and operations of Euler diagrams are triggered spontaneously in problem-solving situations in which users are presented with Euler diagrams. Thus, Euler diagrams appear to have inferential efficacy. This finding comprises the answer to the theoretical question (ii).

My theoretical and empirical findings provide support to the General hypothesis stated in Sections 2 and 1 of Chapter 2 in this thesis:

**Hypothesis on the efficacy of Euler diagrams in syllogistic reasoning.** Euler diagrams can be effective in supporting syllogistic reasoning because they can make explicit the relational information (i.e., inclusion and exclusion relations) contained in categorical sentences and thereby replace

the process of drawing a valid conclusion with the process of manipulating diagrams.

**Hypothesis on the inferential efficacy of diagrams.** Diagrams can be effective in supporting deductive reasoning if the solving processes (i.e. processes of drawing a conclusion) are replaced by processes of manipulating diagrams.

It is also important to consider the question of the cognitive origins of *naturalness* or *effortlessness* in terms of the rules of diagrammatic manipulation; namely, whether diagrammatic manipulations obey natural constraints. This paragraph reviews empirical evidence supporting the view that concrete manipulation of diagrams can be considered a relatively effortless or natural system. We begin by considering the findings of combinatorial manipulations of objects, particularly the nesting cup task. The nesting cup task was developed by Greenfield, Nelson, & Saltzman (1972) to test children's understanding of hierarchical structure, by observing the nesting methods they used with cups of various sizes. Among these methods, the subassembly method is particularly important for understanding the manipulation of Euler diagrams. In the subassembly method, one nesting relation involving two cups is combined with another nesting relation. This method has been reported in young children (Greenfield, Nelson, & Saltzman, 1972; Deloache, Sugarman, & Brown, 1985) and apes (Jonson-Pynn, et al. 1999). It should be noted that manipulations in the nesting cup task affect the relations between several individual objects but also the relations between several clusters of objects. A similar ability is considered to play an important role in the manipulation of unifying the inclusion Euler diagrams, although the identification processes of Euler diagrams are not necessarily based on the concept of size. In general, in diagrammatic representations, diagrammatic objects of various sizes are also identified in terms of named labels. Sloutsky and Lo (1999) reported that, in young children, linguistic labels contribute effectively to similarity judgment among multiple face pictures. These findings suggest that using labels in the identification processes of Euler diagrams would also be expected to be natural and intuitive.

Since it is clear that children can unify diagrams, it is important to determine whether the manipulation of diagrams simulates inference. Experiment 4 in Section 4 of Chapter 3 provided exploratory evidence to address this ques-

tion, by examining children's reasoning with diagrams using a similar method to that in Experiment 1. The results indicated that the differences between Euler and Venn diagrams had a significant effect on the performance of children in the sixth grade (11 years old) in solving valid syllogisms. This finding suggests that even in the case of reasoning by untrained children, the syntactic manipulation of Euler diagrams could be naturally triggered, facilitating the process of combining information from the premises. If children have a certain linguistic ability to understand the structural correspondence between diagrammatic representations and the meaning of a sentence, manipulations of Euler diagrams might thus simulate logical inference, meaning that procedural correspondence holds. As such, diagram manipulations with natural constraint rules appear to constitute an effortless process of reasoning.

Overall, the current findings indicate that reasoning with diagram manipulation can encompass many possible transformational rules based on spatial constraints. It should be noted here that these rules are self-justifiable, in contrast to the usual linguistic deductive reasoning that uses stipulative rules, providing justification externally. It is well known that external justification often leads to a puzzle regarding the justification of a deduction (Dummett, 1973); that is, to justify the deduction, inference rules must be applied. Then, to justify the inference rules, we would need to apply other inference rules, which would continue *ad infinitum* in a circular argument. In contrast, diagrammatic reasoning uses intuitive inference rules based on spatial constraints, so would be expected to prevent circular justification. These conclusions apply to the C1 and C2 rules in the cases of syllogistic reasoning with Euler and linear diagrams. The natural form of logical representation should also be considered here. Given that external diagrams improve interpretation and reasoning performance, mental abstraction is likely to be closely linked with spatial/geometrical properties. Gärdenfors (2000) makes a similar claim in his theory of concept learning, where information is represented by location in geometrical structures of conceptual space. Lakoff and Núñez (2000) investigate cognitive origins of mathematical idea (including the laws of arithmetic, set-theory, and logic, etc) from the viewpoint of *embodied mind*, where the relationship between motor-perceptual control system and human conceptual system is focused. They enable us to view my account of visuo-spatial representation in a much broader context.

Of course, such a diagrammatic system cannot be widely applied to various

situations. Thus, the logic system requires the addition of certain stipulative rules. However, these rules are not controlled by natural or spatial constraints, and in many cases, depend more on symbols than on diagrams. The points above also apply to the inference rule C3, involving the convention of crossing in the NVC and existential cases of syllogistic reasoning with Euler and linear diagrams. This rule may partially break the enclosure of natural constraint rules, moving towards “symbolic logic”, while the application of the rule is a requirement of the simulation of complex reasoning, as in categorical syllogisms.

Regarding the extension with conventional devices, consider the distinctive roles played by two kinds of inference processes underlying deductive reasoning tasks, namely, processes of *proof* (verification, justification) and processes of *dis-proof* (falsification, refutation). A process of proof refers to a process of judging a proposition to be a valid conclusion from a set of premises. In contrast, a process of disproof refers to a process of judging a proposition to be an invalid conclusion. There are two conceptions in symbolic logic, proof-theoretic and model-theoretic, which specify these two processes in different ways. According to the proof-theoretic conception, a process of verification plays a primary role; it is specified as a process of finding a particular formal proof (deduction) in a given proof system, which in turn is specified as a set of axioms and inference rules. Falsification is then characterized as a failure in finding any legitimate proofs. According to the model-theoretic conception, in contrast, the notion of falsification is primitive; it is typically specified as a process of finding a particular counter-model in which all the premises are true but the conclusion is false. Verification is then defined as a failure to find any counter-examples. The two different conceptions, namely, model-theoretic and proof-theoretic, correspond to two research traditions in the cognitive study of deduction, namely the mental-model approach (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991) and the mental logic approach (Rips, 1994; Braine & O’Brien, 1998). The current findings may have implications for the long-standing debate between the two camps concerning the psychological processes underlying deductive reasoning.

As mentioned in Chapter 1, the *Hyperproof* system proposed by Barwise and Etchemendy (1994) is a hybrid interface of logical formulas and diagrams used to support elementary logic teaching. Following standard practice in modern logic, Hyperproof uses proof-theoretic methods for checking validity, and model-theoretic methods in checking invalidity. However, as reported in Stenning, Cox,



and Oberlander (1995), diagrams are less effective in checking invalidity, because it is often manifested as an *indeterminate problem*, that is, a problem in which a given set of premises does not determine a unique answer or solution. Contrary to this influential view, we argue that certain diagrams have efficacy in checking invalidity and explain this property within a novel conception of diagrammatic reasoning, in which both verification and falsification are enabled by a certain type of syntactic manipulation of diagrams. Some of the implications of the current conception for the debate between mental-model versus mental-proof approaches to deductive reasoning are considered below. Specifically, the current exploration of the cognitive properties and functions of the use of diagrams in deductive reasoning can provide novel insights into the nature and competence of sentential inferences with no valid conclusion.

## 5.2 Further discussion: Structural and procedural correspondence

The above-mentioned considerations raise an explanatory question: under what conditions does a diagrammatic representation system exhibit such interpretational or inferential efficacy? There must be a certain type of structural correspondence between a representation and what it represents. Such a relationship can be characterized in various ways: homomorphism (Barwise & Etchemendy, 1991; Barwise & Hammer, 1996), matching relationships (Gurr, Lee, & Stenning, 1998), structural similarity (Gattis, 2004), and so on (see Gurr, 1998; 1999 for a unified classification of the relevant notions of “matching” and “isomorphism”, with various examples). As stated in Chapter 4, the type of structural correspondence is clearly an important factor in determining whether a representational system has *interpretational* efficacy. I propose that structural correspondence exists between the representation systems of Euler diagrams and categorical syllogisms (see Experiments 5-1 and 5-2).

A more pertinent question for my purposes is as follows: what causes a representation system to have *inferential* efficacy (i.e., a type of operational or procedural effectiveness in reasoning processes)? This question has been somewhat neglected in the literature concerning the efficacy of diagrams (cf. the remark by Larkin & Simon, 1987, quoted above). Inferences or derivations are typically analyzed in terms of a *proof system* which is couched within a representational

system. However, I propose that the inferential efficacy of diagrams crucially depends on whether a relationship holds between two proof systems: one for linguistic inferences and one for diagrammatic inferences. Specifically, if what is derivable in one system (e.g., a syllogistic inference system) can faithfully be *simulated* in another system (e.g., a diagrammatic inference system for Euler diagrams) with suitable natural properties, then the latter system would be expected to exert the effectiveness in reasoning processes. In other words, what triggers inferential efficacy is the mapping or correspondence relation, not only between representational elements, but also between also *derivations or proof constructions* built from them. Here, the latter is referred to as “procedural correspondence”. If the correspondence between two inference systems does not hold, the process of checking validity would be unrelated to the process of manipulating diagrams.

Barwise and Shimojima (1995) developed an understanding of diagrammatic reasoning in terms of the notion of *surrogate reasoning*, i.e., reasoning in which inferential processes are partially taken over by external tools such as diagrams. To explain the efficacy of external representations in surrogate reasoning, Shimojima (1996; 1999a) proposed the “constraint (projection or preservation) hypothesis”:

Representations are objects in the world, and as such they obey certain structural constraints that govern their possible formation. The variance in inferential potential of different modes of representation is largely attributable to different ways in which these structural constraints on representations match with the constraints on targets of representation (Shimojima, 1996, pp. 13–14; this was also cited by Shin & Lemon, 2001/2008).

This emphasizes that the efficacy of diagrams in reasoning depends on whether the rules (in his term, “constraints”) of reasoning match those of operations with surrogate objects (diagrams). That is, matching may guarantee the correctness of operating with the diagram in a reasoning task, thus increasing the *reliability* of the operations. However, it should be noted that Shimojima’s explanation of matching in the two representations focuses on the semantic consequence of the assumptions, but not on the proof-theoretical consequence, namely *processes*. The former is concerned with the question of what follows from the assumptions and the latter involves the question of how the consequence follows. In practice,

there might be two inference systems with consequences that are semantically consistent but with proof-theoretical processes that differ. This case could be important for addressing the efficacy of diagrams in inferential processes. Syllogistic reasoning with Euler and Venn diagrams are a typical example of such a case. To clarify the procedural correspondences of two inference systems, the current study examined proof-theoretical analyses of diagrammatic and relevant linguistic reasoning.

Given the concrete task situation, the emphasis of this thesis is on matching between diagrammatic and linguistic systems, rather than matching between a diagram and what it represents. For this reason, clarifying natural interpretations and inferences in the forms of natural languages is within the scope of this study. As such, this thesis proposes a new and extended framework of diagrammatic reasoning, in which diagrams can help to clarify natural inference and interpretation for humans.

Regarding my empirical findings, a *structural correspondence* between the two representation systems is important for a representation system to have interpretational efficacy. This claim is supported by the results of Experiment 5. In that experiment, matching times between syllogistic sentences and Euler diagrams were shorter than those between syllogistic sentences and Venn diagrams. This suggests that the process of extracting relational information from Euler diagrams consists of a single step, whereas Venn diagrams require multiple steps. Venn diagrams may have indirect effects on interpretational efficacy by training users, but Euler diagrams demonstrate additional efficacy (cf. the results in Experiments 1 and 2, where a significant difference of the interpretational effects was reported between Euler diagrams and Venn diagrams). Importantly, when inferential efficacy holds in diagrammatic reasoning, diagram manipulation can act as a *surrogate* linguistic inference (Barwise & Shimojima, 1995). In other words, the diagrammatic inference system can *simulate* the linguistic inference system. This can be examined by the correspondence, not only between representational elements, but also between derivations of multiple inference systems. The Euler diagrams in syllogistic reasoning are well-adapted to the conditions of *procedural correspondence*. Both categorical sentences and Euler diagrams are analyzed as (inclusion and exclusion) relation-based representations, and both sentential syllogistic and Euler diagrammatic inferences are decomposed as inferences with two rules that are based on inclusion and exclusion relations.

## 5.3 Comparison with related literature

In this section, I compare my findings with some related works in logic and cognitive science. Subsection 1 discusses the graphical method of case identification. Subsection 2 discusses the dual process theory of reasoning.

### 5.3.1 Case identification

The representation system of Euler diagrams for syllogisms presented in Stenning and Oberlander (1995) is well known in the field of cognitive science. In Stenning and Oberlander’s system, the existential assumption (an interpretation with existential import, i.e., *Some A are B* is derived from *All A are B*) plays an essential role in the procedure of solving syllogisms, in contrast to the Euler diagrams used in the current study (Mineshima et al., 2008; Mineshima et al., 2012a). To clarify this point, a brief overview of Stenning and Oberlander’s Euler diagrammatic reasoning system is provided below.

Stenning and Oberlander introduced Euler diagrams as a method to identify cases of an individual (type) in a categorical sentence. To take a simple example, in the case of the categorical sentence *Some A are not B*, it follows that an individual that is **a** but not **b**. In the case of a categorical syllogism composed of two premises with three properties —**a**, **b**, and **c**—, one can identify whether an individual has each of the three properties or not (with a total of eight possible states). Note here that an individual must have at least some properties to make a meaningful statement with a categorical sentence. Thus, such case identification is suitable for describing categorical sentences with existential import. In this way, case identification is very similar to the idea underlying mental model theory as described in the previous section. Furthermore, Stenning and Oberlander provided computational algorithms from information about individual cases in two premises to those in the conclusion (i.e., a syllogism-solving model) by identifying individual cases. One implementation is the graphical method using Euler diagrams, and the other is the sentential method by substitution (cf. Stenning & Yule, 1997).

According to their graphical method, a cross ( $\times$ ) is used to indicate the existence of an individual. In a diagram, such a cross inside the region *A* indicates the existence of an individual with the property *A*. Note here that their version of Euler diagrammatic representation differs from that of traditional Euler dia-

grams in Gergonne's system (see Section 3.1 of Chapter 2), and can represent partial information in a single diagram. This is similar to the EUL diagram used in the present study and the mental model representation mentioned in the previous section. As an illustration, consider a syllogism with the premises *No A are B* and *Some B are C*. In Fig. 5.1, the premise *No A are B* is associated with diagram  $D_1^s$  in which one cross is inside  $A$  and another cross is inside  $B$ . The

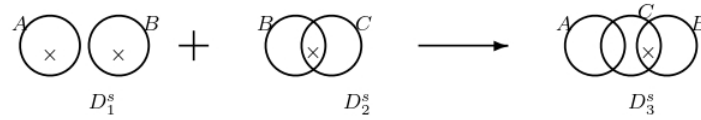


Fig. 5.1 Stenning and Oberlander's Euler diagrammatic reasoning: *No A are B*, *Some B are C*; therefore *Some C are not A*.

premise *Some B are C* is associated with diagram  $D_2^s$  in which the regions of  $B$  and  $C$  are partially overlapping and the cross is inside the intersection of  $B$  and  $C$ . It should be noted that there must be an individual that is  $b$  and  $c$ , but not necessarily an individual that is  $c$  but not  $b$ . By unifying  $D_1^s$  with  $D_2^s$ , we can obtain diagram  $D_3^s$ . Here, the cross in circle  $B$  of  $D_1^s$  is divided into smaller regions and thus disappears, because there is not necessarily an individual that is  $b$  and  $c$ . The cross in circle  $A$  of  $D_1^s$  disappears in same manner. On the other hand, the cross inside the intersection of  $B$  and  $C$  in  $D_2^s$  is not divided and remains unchanged. Here, the cross in  $D_3^s$  means that there is an individual that is not  $a$  but is  $c$ , and is  $b$ . From the diagram, we can extract the correct conclusion "Some  $C$  are not  $A$ ". For another example, consider an invalid syllogism with the premises *All A are B* and *All C are B*, as shown in Fig. 5.2. By unifying

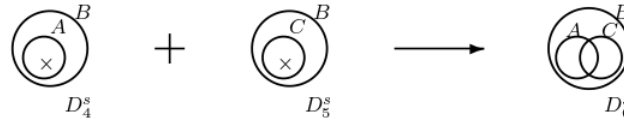


Fig. 5.2 Stenning and Oberlander's Euler diagrammatic reasoning: *All A are B*, *All C are B*; therefore *NVC*.

$D_4^s$  with  $D_5^s$ , the cross in  $D_6^s$  disappears. This means that there is no individual with any properties. From the diagram, we can extract the conclusion "No valid conclusion".

In this way, the existential assumption plays an essential role in the procedure of checking the validity and invalidity of a categorical syllogism, specifically in the process of considering alternative models or situations to verify the sentence. In this respect, there is a close correspondence between their system and mental model theory (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991), as discussed at length in Stenning and Oberlander (1995). Regarding the current findings, the difference between our Euler diagrammatic inference system GDS (Mineshima et al., 2012a) and that of Stenning and Oberlander should be considered. The key difference between the two systems is based on whether existential import is included.

In the current experiments, universal sentences are given non-existential interpretations; but when one prefers the interpretation with existential import, one may adopt a system in which a point is inserted in a relevant region for each universal (affirmative or negative) sentence. Thus, in the case of our Euler diagrams, the issue of existential import is independent, and thus separable, from the account of how the validity of an inference is to be checked. In the case of Stenning and Oberlander’s Euler diagrams, on the other hand, existential import is integrated into the process of validity checking. For instance, Stenning and Oberlander’s diagrams could not account even for NVC cases without focusing on the cross (i.e., whether there is a cross in a specific entity).

Unlike Stenning and Oberlander’s approach, the problem of existential import should be discussed in the literature of pragmatics rather than formal (logical) semantics (cf. Geurts, 2003b; 2007). There are many examples in which existential import does not hold. In using the common expression *All errors are mine*, one is not asserting the existence of an error in their own paper. Thus, an explanation regarding a judgment of the validity or invalidity of an inferences should not depend on pragmatic or contextual inference. Furthermore, Stenning and Oberlander admit the existential import of the predicate term  $B$  in *No  $A$  are  $B$* , as shown in  $D_1^s$  of Fig. 5.1. This represents a more serious problem (cf. Geach, 1962). If one applies the idea to the case *No women are witches*, as Boolos (1984) pointed out, one must acknowledge the existence of witches. Such a problem is known as the square of opposition (e.g., Strawson, 1952) and the existential import of *No  $A$  are  $B$*  is generally unacceptable. Boolos (1984) criticized mental model theory (Johnson-Laird & Bara, 1984) in terms of its existential import assumption for categorical sentences (for a similar criticism from psychology, see

Chater & Oaksford, 1999). The same criticism could be applied to Stenning and Oberlander’s system, which must focus on the cross, expressing the existence of an individual, in not only existential sentences but also non-existential sentences. This indicates that their graphical system is complicated by unnecessary conventional devices in an unhelpful manner.

On the other hand, the previous section indicates that such a process of entertaining alternative possibilities can be formulated, and even constrained, without admitting existential import for universal sentences. For example, in the GDS system, the NVC cases are accounted for as follows: (1) In syllogisms that include existential sentences, invalidity judgments are conducted by the indeterminacy of the relative positional relationship of each point to a region. (2) In syllogisms including non-existential sentences, invalidity judgments are conducted by the indeterminacy of the relation between circles. That is, the graphical system considered so far appears to be free from conventional devices and succeeds in separating the issue of existential import from the core mechanism of validity checking.

### 5.3.2 Dual process theory

My account of the efficacy of diagrams in deductive reasoning can be related to the so-called dual process theory (Evans, 2003; Stanovich & West, 2000). This enables us to view my account from a much broader perspective. In reasoning, the difference between System 1 and 2 is generally studied as the distinction between cognitive processes that are fast, automatic, and unconscious and those that are slow, deliberative, and conscious. In the literature on the psychology of syllogistic reasoning, the dual process has been studied within the framework of belief bias (e.g., Evans, Barston, & Pollard, 1983). However, in the case of syllogisms that take abstract forms as used in my experiments, the dual process is recently argued from the viewpoint of individual differences. In this subsection, discuss on results in light of the findings of dual process theory.

An analysis of the heuristic process provides a starting point. Chater and Oaksford (1999) analyzed the ordering of quantificational sentences in terms of informativeness and showed that the syllogistic inference model, including heuristic rules, is based on the idea of order. To illustrate the model, we will take an example of a syllogism having the premises *All A are B* and *No B are C*. Here, the categorical sentence *All A are B* is more informative than *No B are C*. Fur-

thermore, according to the heuristic rule of *min-heuristic*, in which the least informative premise is selected as the conclusion, *No C are A* is selected more often than *All C are A* as the conclusion. Oaksford and Chater (2001) claimed that such a probabilistic approach provides a computational level theory of System 1 processes. In addition, they reported that many reasoners involve System-1 processes based on heuristics and few reasoners involve System-2 processes based on logical competences such as mental model and mental logic. However, Oaksford and Chater's view of a dual process is in striking contrast to the protocol results of Ford (1994) and Bacon et al. (2003), as stated in Section 1.3 of Chapter 3, in which many reasoners involve System-2 processes. This is considered to be affected by setup of the protocol-specific to the experiment, in which reasoners are asked to make their strategies explicit. In this situation, it could be difficult to solve the syllogistic tasks with an implicitly used heuristic process (such a situation involving the processes with self-explanations may be considered as a special case of *argumentative context*, which was recently proposed in Mercier & Sperber, 2011).

The findings from the above studies indicate that the functioning of the dual process depends on the reasoners' (individuals') attitudes toward deduction tasks. Stenning and Cox (2006) provided comprehensive evidence to support this view of dual process. They divided the participants into rash (corresponding to System 1) and hesitant (corresponding to System 2), as a measure of individual differences in time taken to solve the syllogistic problems. According to their results, the two groups show the different interpretational tendencies of categorical sentences. On the one hand, rash participants take a credulous stance and then make the illicit conversion error of categorical sentence (according to Stenning and van Lambalgen (2008), such a conversion could be explained by non-monotonic closed world reasoning). On the other hand, hesitant participants take a skeptical stance and then (a)symmetry in the relation between subject and predicate in the categorical sentence is detectable. Consequently, the illicit conversion error is blocked. With respect to illicit conversion, we can draw an interesting comparison to my experimental results (Experiment 1, Section 1 of Chapter 3). Many participants in the Linguistic group made the same error as that made by rash-credulous participants. Given the Euler diagrams, namely, in the Euler group, such a conversion error was blocked as in hesitant-skeptical participants. Given the findings on dual process accounts of syllogistic reasoning, we can provide the



following claims about the position of diagrammatic reasoning in dual process accounts: syllogistic tasks in an abstract form should belong to System 2; However, given external diagrams, the tasks could be transposed to concrete manipulations of diagrams which are under the control of System 1. In this sense, we can say that external diagrams change the nature of inference tasks.

Given the appropriate diagrams, various logical reasoning *can* be performed under the control of intuitive system (i.e., System 1). When considered as the study of the proof-theoretical/syntactic aspects of diagram manipulations, diagrammatic reasoning has a potential for providing a bridge between concrete and abstract aspects of reasoning. The fact that elementary logical reasoning can be effortlessly simulated in terms of spatial operations would also be important to understand the cognitive origin of logical ability, i.e., *how* people acquire logical ability in the first place.

## 5.4 Future works

Although the discussion made so far was limited to deductive reasoning with the standard types of quantifiers, i.e., existential and universal quantifiers, it is interesting to see how the current study can be extended to “uncertain” reasoning tasks. One type of uncertain reasoning is “plurative” syllogisms involving quantifiers such as *most* and *few* (cf. Hackl, 2009; Pietroski, Lidz, Hunter & Halberda, 2009; for the cognitive importance of “most” in the connection with generic sentences, see Leslie, Khemlani, & Glucksberg, 2011; Leslie & Gelman, 2012; Khemlani, Leslie, & Glucksberg, 2012). Rescher and Gallagher (1965) reported a conception of diagrammatic representations of the plurative quantifier *most*, which is conventionally represented in Venn diagrams using an arrow, as shown in Fig. 5.3. Furthermore, this diagrammatic method is considered to hold



Fig. 5.3 An Euler diagram with arrow notation for *Most A are B* (left side) and an area proportion Euler diagram (right side)

in the case of conditional reasoning with a probability expression such as *if p then it is very probable that q* (cf. George, 1997; Stevens & Over, 1995). On the other

hand, probability knowledge could also be expressed by the concept of area proportion in logic diagrams such as Euler and Venn diagrams. These are referred to as area proportional Euler/Venn diagrams, and have been intensively investigated in the diagrammatic logic literature (cf. Chow & Ruskey, 2004; Stapleton, Rodgers, & Howse, 2011). For example, in the area proportional diagram shown in Fig. 5.3, the percentage represented by each region is proportional to the area in each region. Such a diagram is expected to be available in probability judgment tasks. To express the base rate condition, for example, Sloman, Over, Slovak, and Stibel (2003) and Yamagishi (2003) used nesting set relation diagrams, which are considered a type of area-proportional Euler diagram. These studies both reported that the use of such a diagram improved participants' performance on a probability judgment task regarding the base rate fallacy.

The relational analysis of quantificational sentences in Section 2 of Chapter 2 is related to the hierarchy of semantic types of various expressions (cf. Montague 1974; Szymanik & Zajenkowski, 2010). For example, the sentence *John is taller than Mary* asserts a (first-order) relation denoted by *is taller than* of two individuals denoted by *John* and *Mary*. Analogously, sentence *All painters are artists* asserts a second-order relation denoted by *all*, i.e., the subset relation of two sets denoted by *painter* and *artist*. Thus we can rank categorical sentences used in syllogistic reasoning tasks in Table 5.1. According to this analysis, categorical

Table 5.1 A classification of types of natural language sentences

Examples	Logical form	Semantic descriptions
<i>John is a painter</i>	$F(a)$	First-order property (a property of an individual)
<i>John is taller than Mary</i>	$R(a, b)$	First-order relation (a relation between individuals)
<i>There is a painter</i>	$Q(F)$	Second-order property (a property of a set)
<i>All painters are artists</i>	$Q(F, G)$	Second-order relation (a relation between sets)

syllogisms may cause complications in two respects: first, they involve *relations* rather than properties, and second, they are concerned not simply with individuals but with a *set* of individuals. Investigations of syllogistic reasoning with sentences involving various levels of relational complexity will also be important in future research.

# Appendix A

## The efficacy of region-based diagrams

Appendix A discusses the efficacy of Venn diagrams. Venn diagrams are characterized as “region-based” diagrams, in sharp contrast to Euler diagrams, i.e., diagrams that are defined in terms of relations. Two experiments are described to examine the interpretational and inferential effects of Venn diagrams. In Section 1, I present an experiment comparing between syllogistic reasoning with Venn diagrams and reasoning with set-theoretical symbolic representations corresponding to Venn diagrams. In Section 2, I present an experiment using Venn diagrams with three circles (3-Venn diagrams) instead of those with two circles (2-Venn diagrams). The results of the two experiments provide support for the claim that Venn diagrams have interpretational efficacy, and that a special version of Venn diagrams, 3-Venn diagram, has inferential efficacy. Based on the fact that Euler diagrams were more effective than 3-Venn diagrams in solving syllogisms, furthermore, I discuss the possibility of a reasoning task (other than traditional categorical syllogisms) which is more appropriately matched to Venn diagrams.

### A.1 Reasoning with Symbolic representations: Experiment 6

The performance of the Venn group in Experiment 1 may be explained by the contribution of Venn diagrams to participants’ interpretations of categorical sentences without playing a substantial role in the inference processes themselves. Hence, people 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. To test this possibility, set-theoretical expressions corresponding to Venn diagrams are introduced, such as  $A \cap \overline{B} = \emptyset$  for “All  $A$  are  $B$ ” and  $A \cap B \neq \emptyset$  for “Some  $A$  are  $B$ ”. It is assumed that these expressions could only provide some information on the meaning of categorical sentences, but could not block specific interpretational errors caused by word order: i.e., conversion errors and figural effects. This experiment is conducted to check whether the difference between the Linguistic and Venn groups was due to the Venn diagrams preventing interpretational errors in categorical sentences. The work of this appendix was obtained by analyzing the data originally collected in Sato, Mineshima, and Takemura (2010b).

### A.1.1 Method

The experiment was conducted in the same manner as Experiment 1, except that in the syllogistic reasoning tasks, Euler and Venn diagrams associated with premises were replaced by corresponding set-theoretical symbolic representations.

#### Participants

Ninety-six undergraduates (mean age  $19.45 \pm 1.01$  SD) in two introductory philosophy classes took part in the experiments, referred to as the “Symbolic group”. Of these, I excluded 32 students: those who left the last three or more questions unanswered (26 students) and those who had participated in my previous pilot experiments (6 students).

#### Materials

**Pretest** Participants in the Symbolic group were presented with 10 set-theoretical expressions listed in Fig. A.1. They were asked to choose the sentences that corresponded with a given representation. The highest possible score on the pretest of the Symbolic group was ten, and the cutoff point was five. The cutoff point was chosen carefully, based upon the results of pilot experiments. Before the pretest, the participants were presented with three examples, as shown in Fig. A.2.

**Syllogistic reasoning tasks** Participants in the Symbolic group were given syllogisms with set-theoretical representations (such as the one in Fig. A.3).

(1)	(2)	(3)	(4)	(5)
$A \cap B \neq \emptyset$	$B \cap \overline{C} \neq \emptyset$	$C \cap B = \emptyset$	$C \cap B \neq \emptyset$	$C \cap \overline{B} = \emptyset$
(6)	(7)	(8)	(9)	(10)
$B \cap \overline{A} \neq \emptyset$	$\overline{B} \cap C \neq \emptyset$	$\overline{A} \cap B = \emptyset$	$\overline{A} \cap \overline{B} = \emptyset$	$B \cap \overline{C} = \emptyset$

Fig. A.1 Set-theoretical symbolic representations used in the pretest

$A \cap \overline{B} = \emptyset$	$B \cap \overline{A} \neq \emptyset$	$A \cap B = \emptyset$
1. All $A$ are $B$ .	1. All $A$ are $B$ .	1. All $A$ are $B$ .
2. No $A$ are $B$ .	2. No $A$ are $B$ .	2. No $A$ are $B$ .
3. Some $A$ are $B$ .	3. Some $A$ are $B$ .	3. Some $A$ are $B$ .
4. Some $A$ are not $B$ .	4. Some $B$ are not $A$ .	4. Some $A$ are not $B$ .
5. None of them.	5. None of them.	5. None of them.
<b>Correct Answer: 1</b>	<b>Correct Answer: 4</b>	<b>Correct Answer: 2</b>

Fig. A.2 The examples in the pretest of the symbolic representations

They were then presented with two premises, and instructed to choose a sentence that corresponded to the correct conclusion. Before the test, the examples in Fig. A.3 were presented.

## Procedure

The experiment was conducted in the same manner as Experiment 1. The instructions for the sentences and set-theoretical representations (the instructions given in Appendix B.1.7) were provided, pretests were conducted, and then the reasoning task was completed.

$$\text{All } B \text{ are } A. \quad B \cap \overline{A} = \emptyset$$

$$\text{All } C \text{ are } B. \quad C \cap \overline{B} = \emptyset$$

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. A.3 An example of the syllogistic reasoning task in the Symbolic group

## A.1.2 Results and Discussion

### Pretest

The accuracy rate for each item in the pretest of set-theoretical symbols was (1) 65.6%, (2) 46.8%, (3) 75.0%, (4) 64.0%, (5) 26.5%, (6) 23.4%, (7) 42.1%, (8) 20.3%, (9) 23.4% (10) 46.8%. In the Symbolic group, 39 students scored less than 5 on the pretest, and were excluded from the final analysis. The correlation coefficient between scores on the pretest and on the syllogistic reasoning tasks in the total symbolic group ( $N = 64$ ) was substantially positive (0.565).

### Syllogistic reasoning tasks

The average accuracy rate for all 31 syllogisms in the Symbolic group was 58.7%. The results for each syllogistic type are shown in Table B.2 of Appendix B.1. The data were compared between the linguistic, Venn, Euler groups in Experiment 1 using one-way ANOVA. The results revealed a significant main effect,  $F(3, 144) = 38.132, p < .001$ . Multiple comparison tests by Ryan's procedure yielded the following results: (i) The accuracy rate of reasoning tasks in the Symbolic group was higher than in the Linguistic group: 46.7% for the Linguistic group and 58.7% for the Symbolic group ( $F(1, 68) = 2.763, p < .05$ ). (ii) There was no significant difference between the accuracy rate of reasoning tasks in the Symbolic group and that in the Venn diagrammatic group: 66.5% for the Venn group and 58.7% for the Symbolic group ( $F(1, 53) = 1.692, p = .10$ ). (iii) The

accuracy rate of reasoning tasks in the Symbolic group was lower than that in the Euler group: 85.2% for the Euler group and 58.7% for the Symbolic group ( $F(1, 68) = 6.123, p < .001$ ). It should be noted that when the participants who failed the pretest were included, a similar pattern of results was found in each comparison: for (iii), there were significant differences,  $p < .001$ ; and for (i) and (ii), there were no significant differences. The accuracy rate for the total Symbolic group (including participants who failed the pretest) was 50.1%.

Regarding the overall accuracy rates of reasoning tasks, there was no significant difference between the Venn group and the Symbolic groups. From this comparison, however, it is unclear whether the use of Venn diagrams blocked the errors caused by misinterpretation such as conversion errors and figural effects. To examine this issue more closely, the 17 invalid syllogisms were divided into four cases based on the types of conclusions that were mistakenly chosen.

**Conversion errors.** 1) In AA2  $all(A, B); all(C, B)$ , AA3  $all(B, A); all(B, C)$  and AA4  $all(A, B); all(B, C)$  syllogisms, 58.5% of the participants in the Linguistic group selected the conclusion “All  $C$  are  $A$ ,” while the rate reduced to 39.9% in the Symbolic group and to 21.1% in the Venn group. These data were also subjected to a one-way ANOVA. There was significant difference between the Linguistic group and the Venn group,  $F(1, 73) = 3.890, p < .001$ . There was no significant difference between the Linguistic group and the Symbolic group,  $F(1, 68) = 1.820, p > .10$ . There was no significant difference between the Symbolic group and the Venn group,  $F(1, 73) = 1.709, p > .10$ .

2) In AI2  $all(A, B); some(C, B)$ , AI4  $all(A, B); some(B, C)$ , IA1  $some(B, A); all(C, B)$  and IA2  $some(A, B); all(C, B)$  syllogisms, 56.1% of the participants in the Linguistic group selected the conclusion “Some  $C$  are  $A$ ,” while the rate reduced to average 41.0% in the Symbolic group and 30.0% in the Venn group. There was significant difference between the Linguistic group and the Venn group,  $F(1, 73) = 2.613, p < .05$ . There was no significant difference between the Linguistic group and the Symbolic group,  $F(1, 68) = 1.429, p > .10$ . There was no significant difference between the Symbolic group and the Venn group,  $F(1, 73) = 0.958, p > .10$ .

3) In AE1  $all(B, A); no(C, B)$ , AE3  $all(B, A); no(B, C)$ , EA3  $no(B, A); all(B, C)$  and EA4  $no(A, B); all(B, C)$  syllogisms, 62.8% of the participants in the Linguistic group selected the conclusion “No  $C$  are  $A$ ,” while the rate reduced to average 30.0% in the Symbolic group and 41.6% in the Venn group. There was a significant difference between the Linguistic group and the Venn group,  $F(1, 73) = 2.214,$

$p < .05$ . There was a significant difference between the Linguistic group and the Symbolic group,  $F(1, 68) = 3.249$ ,  $p < .005$ . There was no significant difference between the Symbolic group and the Venn group,  $F(1, 73) = 1.065$ ,  $p > .10$ .

4) In AO1 *all(B, A); some-not(C, B)*, AO3 *all(B, A); some-not(B, C)*, AO4 *all(A, B); some-not(B, C)*, OA1 *some-not(B, A); all(C, B)*, OA2 *some-not(A, B); all(C, B)* and OA4 *some-not(A, B); all(B, C)* syllogisms, 58.1% of the participants in the Linguistic group selected the conclusion “Some  $C$  are not  $A$ ,” while the rate reduced to an average of 43.3% in the Symbolic group and 28.9% in the Venn group. There was a significant difference between the Linguistic group and the Venn group,  $F(1, 73) = 3.837$ ,  $p < .001$ . There was no significant difference between the Linguistic group and the Symbolic group,  $F(1, 68) = 1.836$ ,  $p > .10$ . There was no significant difference between the Symbolic group and the Venn group,  $F(1, 73) = 1.649$ ,  $p > .10$ .

**Figural effects.** Comparing EI1O and EI4O syllogisms, there was a significant difference between EI1O *no(B, A); some(C, B) : some-not(C, A)* and EI4O *no(A, B); some(B, C) : some-not(C, A)* in the Linguistic group (62.2% for EI1O and 35.6% for EI4O) ( $t(44) = 11.000$ ,  $p < .005$ , t-test, within-subjects design). In contrast, there was no significant difference between EI1O and EI4O in the Symbolic group (48% for EI1O and 48% for EI4O) ( $p = .100$ ). Further, there was no significant difference between EI1O and EI4O in the Venn group (70.0% for EI1O and 73.3% for EI4O) ( $t(29) = 1.000$ ,  $p = .100$ ).

The results revealed that Venn diagrams block all types of interpretational bias, and, in contrast, the effects of symbolic representations are limited to the partial cases: Conversion errors of case (3) and Figural effects. Given these results, even if doubt remains about the interpretational efficacy of set-theoretical symbols, it is clear that Venn diagrams have interpretational efficacy. This suggests that, as far as interpretational errors are concerned, Venn diagrams may have diagram-specific efficacy.

## A.2 Reasoning with Venn diagrams containing three circles: Experiment 7

In Experiment 1, the participants in the Venn group were provided with diagrams consisting of *two* circles that corresponded to the premises of a given syllogism. However, I also test a situation where participants are initially provided with Venn diagrams consisting of *three* circles, or “3-Venn diagrams”, namely  $A$ ,  $B$ ,



and  $C$ , as in  $D_3^v$  and  $D_4^v$  in Fig. A.4 (see also Fig. 2.12 in Chapter 2). 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. To examine the efficacy of 3-Venn diagrams, here I would analyze the experimental data reported in Sato, Mineshima, and Takemura (2010b).

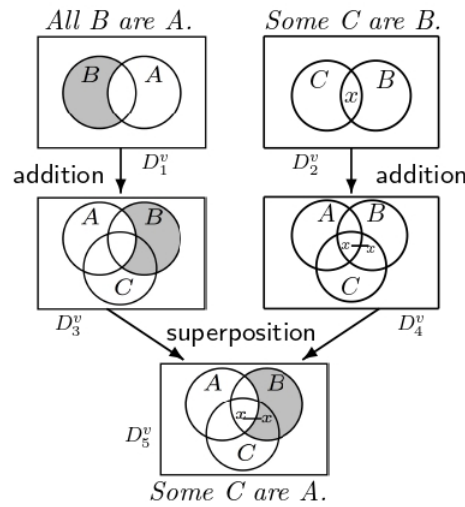


Fig. A.4 Solving a syllogism *All B are A, Some C are B; therefore Some C are A* with Venn diagrams.

## A.2.1 Method

The experiment was conducted in the same manner as Experiment 1, except that in syllogistic reasoning tasks, the Euler and Venn diagrams associated with the premises were replaced by the corresponding 3-Venn diagrams.

### Participants

One hundred and twenty-three undergraduates (mean age  $19.37 \pm 0.95$  SD) in one introductory philosophy class took part in the experiments, referred to as the “3-Venn group”. Of 123 students, I excluded 44; those who left the last three or more questions unanswered (35 students) and those who participated in my pilot experiments (9 students).

## Materials

**Pretest** The participants in the 3-Venn group were presented with the 10 diagrams shown in Fig. A.5. They were asked to choose the sentences that corresponded to a given diagram. The total time taken for the task in the 3-Venn group was 6 minutes (the instructions about the meaning of 3-Venn diagrams took longer than those for the Euler diagrams and Venn diagrams in Experiment 1; for details see Appendix B.1.8). Before the pretest, participants were presented with the three examples shown in Fig. A.6.

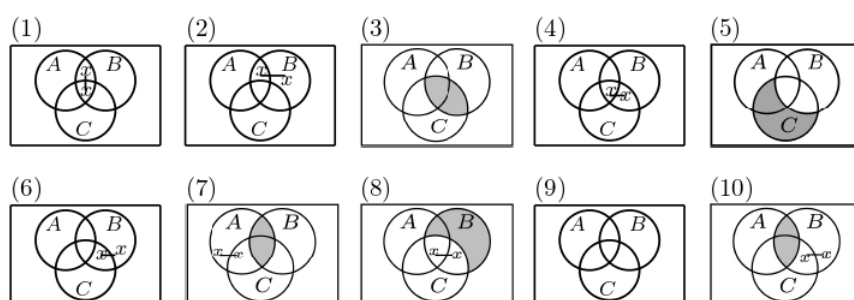
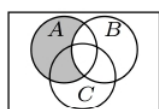
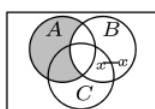


Fig. A.5 3-Venn diagrams used in the pretest



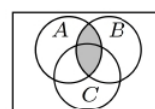
1. All  $A$  are  $B$ .
2. No  $A$  are  $B$ .
3. Some  $A$  are  $B$ .
4. Some  $A$  are not  $B$ .
5. None of them.

**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 and 4

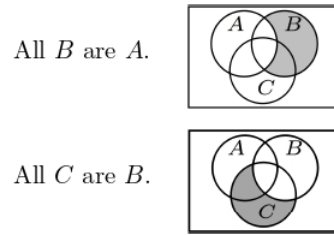


1. All  $A$  are  $B$ .
2. No  $A$  are  $B$ .
3. Some  $A$  are  $B$ .
4. Some  $A$  are not  $B$ .
5. None of them.

**Correct Answer**  
2

Fig. A.6 The examples in the pretest of 3-Venn diagrams

**Syllogistic reasoning tasks** The participants in the 3-Venn group were presented with syllogisms and Venn diagrams involving three circles in their premises (as shown in Fig. A.7). Before the test, the examples in Fig. A.7 were presented.



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. A.7 Example of reasoning task of 3-Venn group

## Procedure

The experiment was conducted in the same manner as Experiment 1. The instructions of sentences and diagrams were provided, pretests were conducted, and the reasoning task was undertaken. Unlike other groups (in Experiment 1), the 3-Venn group was given 3 minutes to read two pages of instructions about the meaning of the diagrams. These time limits were set based on the results of my pilot experiments.

## A.2.2 Results and Discussion

### Pretest

The accuracy rate of each item in the pretest of 3-Venn diagrams was (1) 79.7%, (2) 50.6%, (3) 79.7%, (4) 79.7%, (5) 48.1%, (6) 70.9%, (7) 31.6%, (8) 34.2%, (9) 81.0% (10) 67.1%. In the 3-Venn group, 41 students scored less than 8 on the pretest, and were excluded from the following analysis. The correlation coefficient between scores on the pretest and on the syllogistic reasoning tasks in the total 3-Venn diagrammatic group ( $N = 79$ ) was substantially positive (0.606).

## Syllogistic reasoning tasks

The average accuracy rates across all 31 syllogisms in the 3-Venn group was 75.7%. The results for each syllogistic type are shown in Table B.2 of Appendix B.1. The data were compared with the Linguistic, Venn, Euler groups in Experiment 1 using one-way ANOVA. There was a significant main effect,  $F(3, 157) = 35.204, p < .001$ . Multiple comparison tests using Ryan's procedure yielded the following results: (i) The accuracy rate of reasoning tasks in the 3-Venn group was higher than in the Linguistic group: 46.7% for the Linguistic group and 75.7% for the 3-Venn group ( $F(1, 81) = 7.126, p < .001$ ). (ii) The accuracy rate of reasoning tasks in the 3-Venn group was higher than that in the Linguistic group: 66.5% for the Venn group and 75.7% for the 3-Venn group ( $F(1, 66) = 2.008$ , at a reduced threshold  $p < .10$ ). (iii) The accuracy rate of reasoning tasks in the 3-Venn group was lower than in the Euler group: 85.2% for the Euler group and 58.7% for the 3-Venn group ( $F(1, 81) = 2.338, p < .05$ ). It should be noted that when participants who failed the pretest were included, a similar pattern of results was found in each comparison: for (i), there were significant differences,  $p < .01$ ; for (iii), there were significant differences,  $p < .001$ ; for (ii), there were no significant differences. The overall accuracy rate for the total 3-Venn group (including those who failed the pretest) was 59.4%.

Syllogistic reasoning performance in the 3-Venn group was significantly better than that in the Venn group. This result supports my prediction that 3-Venn diagrams have inferential efficacy, suggesting that they could play a substantial role in reasoning processes themselves. However, the performance of the 3-Venn group was significantly worse than that in the Euler group. To explain this difference, more detailed performance data from the 3-Venn group are shown in Table A.1, in which syllogisms are classified into four cases: non-existential cases with a valid conclusion, existential cases with a valid conclusion, non-existential cases with no valid conclusion (NVC), and existential cases with NVC.

The data were analyzed using a two-way ANOVA. There was a significant main effect of the between-subjects factor, ( $F(1, 81) = 19.136, p < .001$ ). There was a significant main effect for the within-subjects factor, ( $F(3, 243) = 11.457, p < .001$ ). There was a significant interaction effect between the two factors, ( $F(3, 243) = 13.034, p < .001$ ). Multiple comparison tests were conducted using Ryan's procedure. (i) The performance in non-existential cases with a valid conclusion in the 3-Venn group was significantly worse than in the Euler group,  $F(1, 324) =$

Table A.1 The average accuracy rates of the four cases in the 3-Venn group (compared with the Euler group)

	Valid Conclusion		NVC	
	<i>non-existential</i>	<i>existential</i>	<i>non-existential</i>	<i>existential</i>
3-Venn group	74.7%	86.5%	73.6%	53.9%
Euler group	97.2%	81.5%	86.2%	85.2%

19.228,  $p < .001$ . (ii) There was no significant difference between the performance in existential cases with valid conclusions in the 3-Venn group and in the Euler group. (iii) Performance in non-existential cases with NVC in the 3-Venn group was significantly worse than in the Euler group,  $F(1, 324) = 6.049$ ,  $p < .05$ . (iv) Performance in existential cases with NVC in the 3-Venn group was significantly worse than in the Euler group,  $F(1, 324) = 37.330$ ,  $p < .001$ .

These results indicate that 3-Venn diagrams are less tractable than Euler diagrams in non-existential cases with valid conclusions, non-existential cases with NVC, and existential cases with NVC. Here, it is important to examine the possible solving (superposition) processes of syllogisms using 3-Venn diagrams, corresponding to the above four cases, as shown in Figs A.8, A.9, A.10, and A.11. However, it is observed here that each superposition processes are relatively easy to handle. The difficulty in syllogism solving with 3-Venn diagrams might be attributed to the difficulty in the process of drawing a conclusion from an internally constructed diagram. Such a process of extracting information may be formulated as a process of *deletion*. A deletion step in Euler diagrammatic reasoning (as shown on the left in Fig. A.12) is simple in that it only requires the removal of a circle without adjusting any other part of the diagram. In contrast, deletion in 3-Venn diagrammatic reasoning is relatively complicated. In particular, in the step from  $D_3$  to  $D_4$  in Fig. A.13, which is an instance of invalid syllogism (the part of process of Fig. A.10), one has to remove a circle and shading, assemble the shading information in several minimal regions, then judge the relationship between  $A$  and  $C$ . The same can be said of non-existential cases with valid conclusions such as Fig A.8, and existential cases with NVC such as A.11. Such complexities in deletion steps may reflect complexities of the processes of observing conclusions, and hence cause difficulty in this type of syllogism.

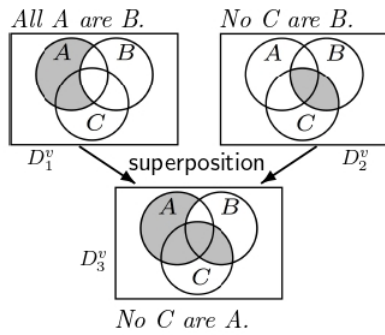


Fig. A.8 Solving a non-existential syllogism with valid conclusion using 3-Venn diagrams.

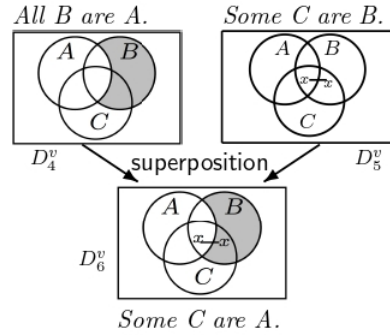


Fig. A.9 Solving an existential syllogism with valid conclusion using 3-Venn diagrams.

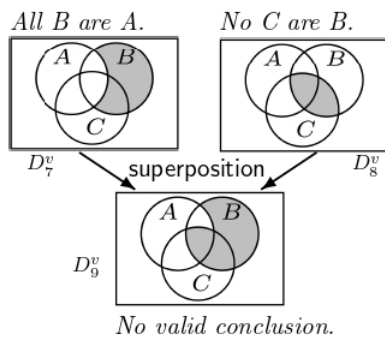


Fig. A.10 Solving a non-existential syllogism with NVC using 3-Venn diagrams.

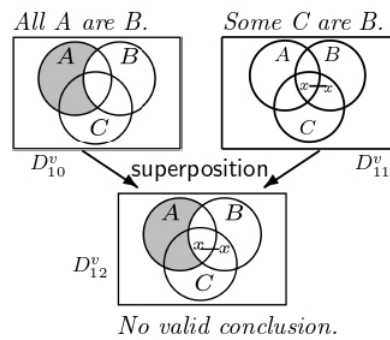


Fig. A.11 Solving an existential syllogism with NVC using 3-Venn diagrams.

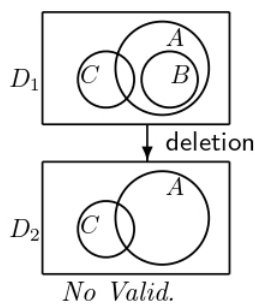


Fig. A.12 Deletion step in Euler diagrams.

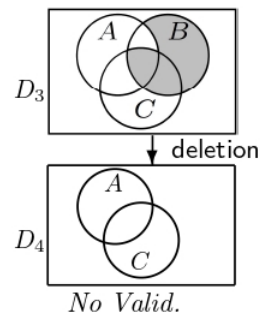


Fig. A.13 Deletion step in Venn diagrams.

In contrast, existential cases with a valid conclusion, such as that shown in Fig A.9, appear to be relatively easy. In this type of case, users only need to recognize the relationship between the point  $x$  and a relevant circle, and the above complexities related to minimal regions can be omitted.

Venn diagrams have a fixed configuration of circles and represent set relationships *indirectly*, by stipulating that shaded regions denote an empty set. Accordingly, the process of extracting relevant relational information from a Venn diagram would be expected to proceed over several steps. As a concrete example, consider how the reasoner could extract the correct relational representation from the Venn diagram  $D_3^v$  in Fig. A.8. Let us refer to a region inside some circles and outside the rest of the circles (possibly none) in a diagram a *minimal* region. Thus the diagram  $D_3^v$  has seven minimal regions. First, the reasoner needs to check whether each minimal region is shaded, as the lower-level information. In this example, four regions  $(A, \overline{B}, \overline{C})$ ,  $(A, \overline{B}, C)$ ,  $(A, B, C)$ ,  $(\overline{A}, B, C)$  are shaded. Next, the reasoner internally builds the *segments* that join the shaded minimal regions continuously to each other, as well as those that join the unshaded minimal regions. This step enables the reasoner to receive higher-level information from the diagram: “there is nothing which is  $C$  and  $A$ ”. The reasoner would then be able to paraphrase this information as “No  $C$  are  $A$ ”, which corresponds to the required information in a syllogistic inference. Such complexities could be considered to cause difficulties in extracting the required information from 3-Venn diagrams.

In Experiment 5, I examined matching time performance between a categorical sentence and a corresponding diagram (or vice versa). When diagrams do not contain a point  $x$  (i.e., the case of non-existential sentences), the response time to match a syllogistic sentence with a corresponding Venn diagram (or vice versa) was longer than the response time in the case of Euler diagrams. To interpret the present results, consider the results in Table A.1 in Experiment 7, showing that non-existential syllogisms were intractable cases and existential syllogisms with valid conclusions were relatively tractable cases. Here we can observe the correspondence between tractable/intractable cases of information extraction and tractable/intractable cases of syllogistic reasoning. Thus, the present results, particularly the connection to those of Experiment 5-2, provide evidence for the hypothesis on the structural correspondence of interpretation, described in Section 2 of Chapter 4 and Section 2 of Chapter 6. These experimental findings

suggest that, in the case of 3-Venn diagrams, the correspondence with the categorical sentences holds indirectly as the result of the complex integration of several minimal regions; that is, the correspondence between representational elements does not hold. Thus, in the overall inference process, including information extraction, manipulation of 3-Venn diagrams does not necessarily simulate syllogistic inference; that is, the procedural correspondence does not hold. This finding is in contrast to the case with Euler diagrams.

In the representational system of Venn diagrams, configurations of circles are as fixed as the maximal regions among circles. Meaningful information is expressed in the system by the addition of syntactic devices, such as shading and points, to each region. The same is true of matrix (or tabular) representation, and Venn diagrams are sometimes regarded as a special case of matrix representation, referred to as a Karnaugh map (Karnaugh, 1953; Wickes, 1968). In matrix representations, in general, a cell is constructed at the intersection of two variables, and meaningful information is expressed by adding a number or mark to each cell. Novick and Hurley (2001) and Novick (2006) analyzed the structural properties of three spatial diagrams: hierarchies, matrices, and networks. Based on the participants' selection of the diagrams in several situations, they showed that while static information is appropriately represented by matrices, networks are more suitable for dynamic information involving the movement of diagrammatic components. Given these findings, it is clear that 2-Venn diagrams (rather than 3-Venn diagrams) are not suitable for categorical syllogistic reasoning that adds a new term in one premise to two terms in the other premise. On the other hand, Euler diagrams and linear diagrams are obviously more suitable for the diagram manipulations necessary for categorical syllogisms, because they can be transformed into a graph, as in a network (cf. Mineshima et al. 2012a).

On the other hand, we could expect Venn diagrams to have efficacy in the case of deductive reasoning with one premise containing the required full terms. The following studies of propositional reasoning provide some initial support for this approach. In several early studies, Schwartz (1971) and Schwartz and Fattaleh (1972) reported that matrix representations are more effective than sentential representations in solving problems concerned with logical connectives such as conjunction and disjunction. In the literature on user interfaces for information retrieval systems, much research has focused on graphical interfaces, including Venn-like diagrams. In keyword searching, people use one or more keywords, then



formulate queries by combining the keywords with Boolean expressions such as AND/OR/NOT. Several evaluative studies reported by Michard (1982), Hertzum and Frøkjær (1996) and Jones, McInnes and Staveley (1999) have indicated that Venn like diagrams can be effective metaphors to help users interpret Boolean expressions. These studies provide partial support for the current proposal regarding the efficacy of Venn diagrams.

# Appendix B

## Supplemental data of the experiments

### B.1 Instructions used in the experiments

#### B.1.1 Instructions on the meaning of categorical statements

In this experiment, you will solve a reasoning task. The meanings of the sentences used in this experiment are defined as follows.

- (1) “All  $A$  are  $B$ ” means that if there are objects which are  $A$ , all of them are  $B$ .
- (2) “No  $A$  are  $B$ ” means that if there are objects which are  $A$ , none of them are  $B$ .
- (3) “Some  $A$  are  $B$ ” means that there are some objects which are  $A$  and  $B$ .
- (4) “Some  $A$  are not  $B$ ” means that there are some objects which are  $A$  but not  $B$ .

#### **NB**

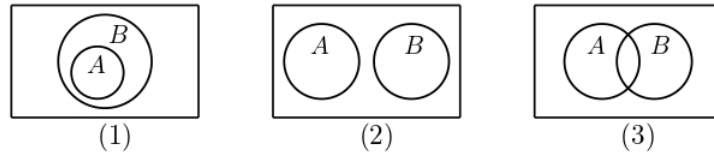
“All  $A$  are  $B$ ” does not imply that there are some objects which are  $A$ . Thus, “All  $A$  are  $B$ ” does not imply “Some  $A$  are  $B$ ”.

“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  $B$ ”.

### B.1.2 Instructions on the meaning of Euler (EUL) diagrams

You may use diagrams in solving reasoning tasks. The meaning of a diagram used in this experiment is defined as follows.

**A circle is used to denote a set of objects.**



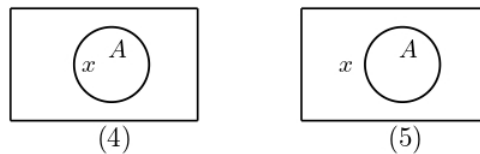
- In Diagram (1) the region of  $A$  is inside of the region of  $B$ . This means that if there are objects which are  $A$ , all of them are  $B$ .
- In Diagram (2), the region of  $A$  is outside of the region of  $B$ . This means that there are no objects which are  $A$  and  $B$ .
- In Diagram (3), the regions of  $A$  and  $B$  partly overlap each other. This means that the relationship between the set of objects which are  $A$  and the set of objects which are  $B$  is unknown.

It should be noted that diagram (3) says nothing about whether there are some objects which are both  $A$  and  $B$ .

**NB**

Diagrams (1)–(3) do not mean that there are some objects which are  $A$  and/or  $B$ . These diagrams say nothing about the existence of objects.

**Point  $x$  is used to indicate the existence of an object.**

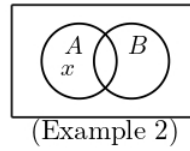
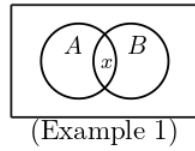


- In Diagram (4), point  $x$  is inside of the region of  $A$ . This means that there is an object which is  $A$ .
- In Diagram (5), point  $x$  is outside of the region of  $A$ . This means that there is an object which is not  $A$ .

**NB**

It is unknown whether there is an object in a region where point  $x$  is absent. For example, diagram (4) says nothing about whether there is an object which **is not**  $A$ . Similarly, diagram (5) says nothing about whether there is an object which **is**  $A$ .

By combining diagrams (1)-(5), we can compose more complex diagrams. We give some examples:



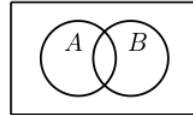
- In Example 1,  $x$  is inside of the intersection of  $A$  and  $B$ . This means that there is an object which is  $A$  and  $B$ .
- In Example 2,  $x$  is inside of  $A$  but outside of  $B$ . This means that there is an object which is  $A$  but not  $B$ .

(It should be noted that Example 2 says nothing about whether there is an object which is both  $A$  and  $B$ .)

### B.1.3 Instructions on the meaning of Venn diagrams

You may use diagrams in solving reasoning tasks. The meaning of a diagram used in this experiment is defined as follows.

**A circle is used to denote a set of objects.**



primary diagram

- In the primary diagram, the regions of  $A$  and  $B$  partly overlap each other. This means that the relationship between the set of objects which are  $A$  and the set of objects which are  $B$  is unknown.

(It should be noted that the primary diagram says nothing about whether there are some objects which are both  $A$  and  $B$ .)

**NB**

The primary diagram does not mean that there are some objects which are  $A$  and/or  $B$ . These diagrams say nothing about the existence of objects.

**Point  $x$  is used to indicate the existence of an object.**



(i)



(ii)

- By putting point  $x$  inside of the region of  $A$ , (i) means that there is an object which is  $A$ .
- By putting point  $x$  outside of the region of  $A$ , (ii) means that there is an object which is not  $A$ .

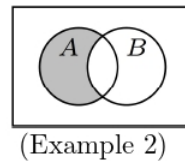
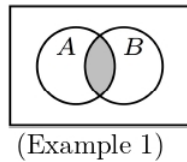
**NB**

It is unknown whether there is an object in a region where point  $x$  is absent. For example, (i) says nothing about whether there is an object which **is not**  $A$ . Similarly, (ii) says nothing about whether there is an object which **is**  $A$ .

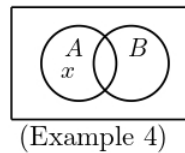
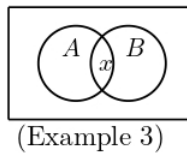
By shading inside of the region of  $A$ , (iii) means that there is nothing which is  $A$ .



By combining primary diagram and (i)-(iii), we can compose more complex diagrams. We give some examples:



- In Example 1, shading is inside of the intersection of  $A$  and  $B$ . This means that there is nothing which is  $A$  and  $B$ .
- In Example 2, shading is inside of  $A$  but outside of  $B$ . This means that there is nothing which is  $A$  but not  $B$ ; this means that if there are objects which are  $A$ , all of them are  $B$ .



- In Example 3,  $x$  is inside of the intersection of  $A$  and  $B$ . This means that there is an object which is  $A$  and  $B$ .
- In Example 4,  $x$  is inside of  $A$  but outside of  $B$ . This means that there is an object which is  $A$  but not  $B$ .

It should be noted that Example 4 says nothing about whether there is an object which is both  $A$  and  $B$ .

### B.1.4 Instructions on the meaning of Linear diagrams

You may use diagrams in solving reasoning tasks. The meaning of a diagram used in this experiment is defined as follows.

**The line  $A$  is used to denote a set of objects of  $A$ .**

$\underline{\hspace{2cm}}A$

$\underline{\hspace{1cm}}\underline{\hspace{1cm}}A\underline{\hspace{1cm}}B$        $\underline{\hspace{1cm}}A \quad \underline{\hspace{1cm}}B$        $\underline{\hspace{1cm}}\underline{\hspace{1cm}}A \quad \underline{\hspace{1cm}}B$   
 (1)                                      (2)                                      (3)

- In Diagram (1), the line  $A$  is a part of the line  $B$ . This means that if there are objects which are  $A$ , all of them are  $B$  (for simplicity, the line  $A$  and the line  $B$  are displaced from same position).
- In Diagram (2), the line  $A$  is outside of the line  $B$ . This means that there are no objects which are  $A$  and  $B$ .
- In Diagram (3), the lines  $A$  and  $B$  partly overlap each other. This means that the relationship between the set of objects which are  $A$  and the set of objects which are  $B$  is unknown.

#### NB

Diagrams (1)–(3) do not mean that there are some objects which are  $A$  and/or  $B$ . These diagrams say nothing about the existence of objects.

**Point  $x$  is used to indicate the existence of an object.**

$\underline{\hspace{1cm}}\blacksquare^x\underline{\hspace{1cm}}A$                                        $\blacksquare^x \quad \underline{\hspace{1cm}}A$   
 (4)    (5)

- In Diagram (4), point  $x$  is on the line  $A$ . This means that there is an object which is  $A$ .
- Diagram (5), point  $x$  is outside of the line  $A$ . This means that there is an object which is not  $A$ .

#### NB

For example, Diagram (4) says nothing about whether there is an object which **is not**  $A$ . Similarly, Diagram (5) says nothing about whether there is an object which **is**  $A$ .

By combining diagrams (1)-(5), we can compose more complex diagrams. We give some examples:



- In Example 1,  $x$  is inside of the intersection of  $A$  and  $B$ . This means that there is an object which is  $A$  and  $B$ .
- In Example 2,  $x$  is inside of  $A$  but outside of  $B$ . This means that there is an object which is  $A$  but not  $B$ .

(It should be noted that Example 2 says nothing about whether there is an object which is both  $A$  and  $B$ .)

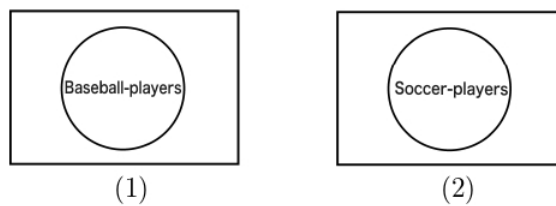


### B.1.5 Instructions on the meaning of Euler diagrams (for children)

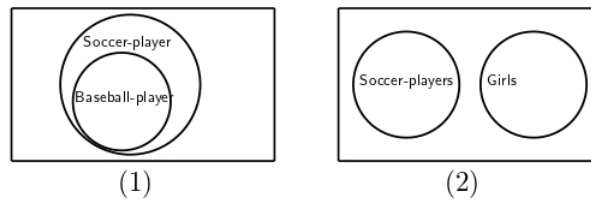
You may use the diagrams in solving reasoning tasks. The meaning of a diagram used in this experiment is defined as follows.

The circle is used to denote the set of people.

Circle (1) denotes the set of **baseball-players** and Circle (2) denotes the set of **soccer-players**.



By combining two circles, we can represent several cases.



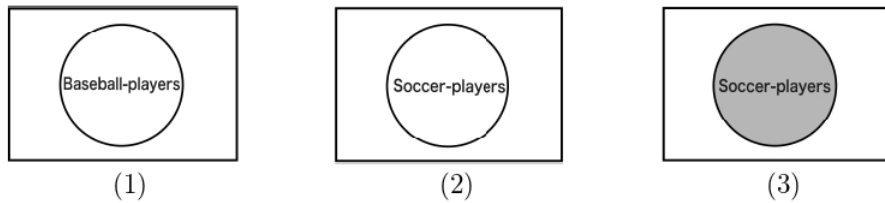
- In Diagram (1), the region of **Baseball-players** is inside of the region of **Soccer-players**. This means that all **Baseball-players** are **Soccer-players**.
- In Diagram (2), the region of **Soccer-players** is outside of the region of **Girls**. This means that there is no person which is a **Soccer-player** and a **Girls**; this mean that No **Soccer-player** are **Girls**.

### B.1.6 Instructions on the meaning of Venn diagrams (for children)

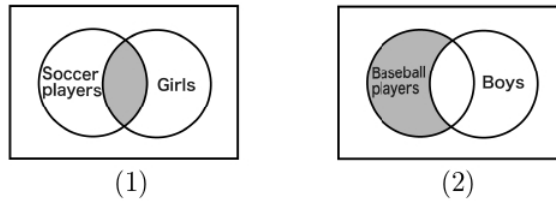
You may use the diagrams in solving reasoning tasks. The meaning of a diagram used in this experiment is defined as follows.

A circle is used to denote a set of people.

Circle (1) denotes the set of **baseball-players** and Circle (2) denotes the set of **soccer-players**. Furthermore, by shading inside of region of **soccer-players**, Circle (3) means that there is no person who is a **soccer-player**.



By combining two circles, we can represent several cases.



- In Diagram (1), shading is inside of intersection of **Soccer-player** and **Girls**. This means that there is no person who is a **Soccer-player** and a **Girl**; this means that **No Soccer-players are Girls**.
- In Diagram (2), shading is inside of **Baseball-players** but outside of **Boys**. This means that there is no person who is a **Baseball-player** but not a **Boys**; this means that **all Baseball-players are Boys**.

### B.1.7 Instructions on the meaning of Set-theoretical representations

You may use symbols in solving reasoning tasks. The meaning of a symbol used in this experiment is defined as follows.

Collections of objects are called “sets”. Individual objects contained in a set are called “elements”.

Capital letters of alphabet ( $A, B, C, \dots$ ) are used to denote sets. If the set  $A$  is the uneven numbers that are less than 10, for example, the elements of  $A$  are as follows: 1, 3, 5, 7, 9

$$\overline{A}$$

(i)

$$A \cap B$$

(ii)

- Convention (i), which is a complement of the set  $A$ , indicates a set of objects that are not contained in  $A$ .
- Convention (ii), which is an intersection of the sets  $A$  and  $B$ , indicates the set of all objects that are elements of both  $A$  and  $B$ . For example, if  $A$  is the set of women and  $B$  is the set of students,  $A \cap B$  is the set of people which are students and women.

$$A = \emptyset$$

(iii)

$$A \neq \emptyset$$

(iv)

- Convention (iii), in which the set  $A$  is an empty set, indicates that set  $A$  has no elements.
- Convention (iv) indicates that the set  $A$  is not empty set; this means that there is an element of  $A$ .

By combining conventions (i)-(iv), we can compose several relationships between sets. We give some examples:

$$A \cap B = \emptyset \qquad A \cap \overline{B} = \emptyset$$

(Example 1) \qquad (Example 2)

- Example 1, in which the intersection of  $A$  and  $B$  is an empty set, means that there is nothing which is  $A$  and  $B$ ; this means that if there are objects which are  $A$ , none of them are  $B$ .
- Example 2, in which the intersection between  $A$  and the complement of  $B$  is an empty set, means that there is nothing which is  $A$  and not  $B$ ; this means that if there are objects which are  $A$ , all of them are  $B$ .

#### NB

Examples 1 and 2 say nothing about whether there is an object that is not  $A$ .

$$A \cap B \neq \emptyset \qquad A \cap \overline{B} \neq \emptyset$$

(Example 3) \qquad (Example 4)

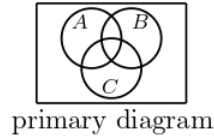
- Example 3, in which the intersection of  $A$  and  $B$  is not an empty set, means that there is an object which is  $A$  and  $B$ .
- Example 4, in which the intersection between  $A$  and the complement of  $B$  is not an empty set, means that there is an object which is  $A$  and not  $B$ .

Examples 3 and 4 indicate that there are some objects which are  $A$ .

### B.1.8 Instructions on the meaning of 3-Venn diagrams

You may use diagrams in solving reasoning tasks. The meaning of a diagram used in this experiment is defined as follows.

**A circle is used to denote a set of objects.**

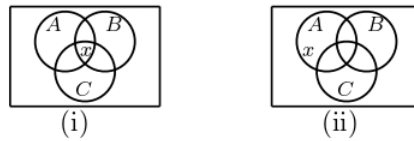


- In the primary diagram, the regions of  $A$ ,  $B$ , and  $C$  partly overlap each other. This means that the relationship among the set of objects which are  $A$ ,  $B$ , and  $C$  is unknown.

**NB**

The primary diagram says nothing about the existence of objects.

**Point  $x$  is used to indicate the existence of an object.**

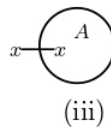


- By putting point  $x$  inside of the intersection of  $A$ ,  $B$ , and  $C$ , (i) means that there is an object which is  $A$ ,  $B$ , and  $C$ .
- By putting point  $x$  inside of  $A$  but outside of the regions of  $B$  and  $C$ , (ii) means that there is an object which is  $A$  but not  $B$  and  $C$ .

**NB**

It is unknown whether there is an object in a region where point  $x$  is absent. For example, (ii) says nothing about whether there is an object which **is**  $A$ ,  $B$ , and  $C$ .

**Linking two points in different regions is used to indicate the existence of an object in either of those regions.**



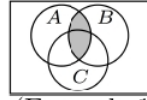
- By linking between the point  $x$  inside of  $A$  and outside of  $B$ , (iii) means that either there is an object which is  $A$  or there is an object which is not  $A$ .



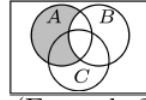
(iv)

- By shading the inside of the region of  $A$ , (iv) means that there is nothing which is  $A$ .

By combining the primary diagram and (i)-(iv), we can compose more complex diagrams. We give some examples:

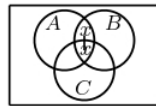


(Example 1)

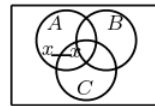


(Example 2)

- In Example 1, shading is inside of the intersection of  $A$  and  $B$ . This means that there is nothing which is  $A$  and  $B$ . Note here that this diagram says nothing about the relationship between  $A$  and  $C$ , or between  $B$  and  $C$ .
- In Example 2, shading is inside of  $A$  but outside of  $B$ . This means that there is nothing which is  $A$  but not  $B$ ; this means that if there are objects which are  $A$ , all of them are  $B$ . Note here that this diagram says nothing about the relationship between  $A$  and  $C$ , or between  $B$  and  $C$ .



(Example 3)



(Example 4)

- In Example 3,  $x$  is inside of the intersection of  $A$  and  $B$  and are linked to each other. This means that there is an object which is  $A$  and  $B$ . Note that it is unknown whether the object is  $C$  or not.
- In Example 4,  $x$  is inside of  $A$  but outside of  $B$  and linked each other. This means that there is an object which is  $A$  but not  $B$ . Note that it is unknown whether the object is  $C$  or not.

It should be noted that Example 4 cannot determine whether there is an object which is both  $A$  and  $B$ .

## **B.2 The results of each syllogistic type in the six groups**

The results of each syllogistic type in the Linguistic, Venn, and Euler groups are shown in Table B.1. The results of each syllogistic type in the Linear, Symbolic, and 3-Venn groups are shown in Table B.2. Numbers indicate the percentage of total responses to each syllogism. Bold type refers to a valid conclusion by the standard of predicate logic. For simplicity, we excluded the conclusions of the so-called “weak” syllogisms (i.e., syllogisms whose validity depends on the existential import of subject term) from valid answers.

Table B.1 Response distributions for 31 syllogisms in the Linguistic group, Venn group and Euler group (Bold type indicates a valid conclusion)

code $\mathcal{C}$ figure	premises 1st, 2nd	Linguistic group, N=45					Venn group, N=30					Euler group, N=45				
		conclusion*					conclusion*					conclusion*				
		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	<b>31.1</b>	23.3	0.0	3.3	0.0	<b>73.3</b>	4.4	0.0	0.0	0.0	<b>95.6</b>
AA3	$all(B, A); all(B, C)$	60.0	0.0	26.7	0.0	<b>13.3</b>	16.6	0.0	16.6	0.0	<b>63.3</b>	0.0	0.0	13.3	0.0	<b>86.7</b>
AA4	$all(A, B); all(B, C)$	60.0	0.0	28.9	0.0	<b>11.1</b>	23.3	3.3	16.6	3.3	<b>53.3</b>	8.9	0.0	11.1	2.2	<b>77.8</b>
AI1	$all(B, A); some(C, B)$	2.2	2.2	<b>88.9</b>	2.2	4.4	0.0	0.0	<b>93.3</b>	0.0	6.6	0.0	0.0	<b>100.0</b>	0.0	0.0
AI2	$all(A, B); some(C, B)$	0.0	0.0	55.6	17.8	<b>26.7</b>	0.0	3.3	30.0	0.0	<b>66.6</b>	0.0	0.0	6.7	2.2	<b>88.9</b>
AI3	$all(B, A); some(B, C)$	4.4	2.2	<b>80.0</b>	6.7	6.7	0.0	0.0	<b>80.0</b>	6.6	13.3	0.0	0.0	<b>84.4</b>	0.0	15.6
AI4	$all(A, B); some(B, C)$	4.4	0.0	57.8	6.7	<b>31.1</b>	0.0	0.0	33.3	0.0	<b>66.7</b>	0.0	0.0	15.6	0.0	<b>84.4</b>
IA1	$some(B, A); all(C, B)$	2.2	2.2	60.0	8.9	<b>26.7</b>	3.3	0.0	33.3	0.0	<b>63.3</b>	0.0	0.0	8.9	2.2	<b>84.4</b>
IA2	$some(A, B); all(C, B)$	6.7	0.0	51.1	11.1	<b>31.1</b>	3.3	3.3	23.3	0.0	<b>70.0</b>	0.0	0.0	13.3	2.2	<b>84.4</b>
IA3	$some(B, A); all(B, C)$	0.0	2.2	<b>93.3</b>	0.0	4.4	0.0	0.0	<b>66.7</b>	0.0	33.3	0.0	0.0	<b>75.6</b>	0.0	24.4
IA4	$some(A, B); all(B, C)$	11.1	0.0	<b>73.3</b>	6.7	6.7	0.0	3.3	<b>50.0</b>	3.3	43.3	0.0	0.0	<b>68.9</b>	0.0	31.1
AE1	$all(B, A); no(C, B)$	0.0	64.4	0.0	6.7	<b>26.7</b>	0.0	33.3	3.3	3.3	<b>60.0</b>	0.0	6.7	0.0	4.4	<b>88.9</b>
AE2	$all(A, B); no(C, B)$	2.2	<b>93.3</b>	0.0	2.2	2.2	0.0	<b>86.7</b>	0.0	6.6	6.6	0.0	<b>97.8</b>	0.0	0.0	2.2
AE3	$all(B, A); no(B, C)$	0.0	64.4	2.2	11.1	<b>20.0</b>	3.3	36.6	0.0	10.0	<b>46.6</b>	0.0	15.6	0.0	4.4	<b>80.0</b>
AE4	$all(A, B); no(B, C)$	0.0	<b>77.8</b>	4.4	6.7	11.1	0.0	<b>80.0</b>	0.0	6.6	10.0	0.0	<b>95.6</b>	0.0	2.2	2.2
EA1	$no(B, A); all(C, B)$	0.0	<b>91.1</b>	0.0	2.2	2.2	0.0	<b>96.6</b>	0.0	0.0	3.3	0.0	<b>97.8</b>	0.0	0.0	2.2
EA2	$no(A, B); all(C, B)$	2.2	<b>88.9</b>	0.0	4.4	4.4	3.3	<b>90.0</b>	3.3	3.3	0.0	0.0	<b>100.0</b>	0.0	0.0	0.0
EA3	$no(B, A); all(B, C)$	0.0	62.2	0.0	20.0	<b>17.8</b>	0.0	53.3	0.0	10.0	<b>36.6</b>	0.0	11.1	0.0	4.4	<b>84.4</b>
EA4	$no(A, B); all(B, C)$	2.2	60.0	2.2	13.3	<b>17.8</b>	0.0	43.3	3.3	10.0	<b>36.6</b>	0.0	6.7	0.0	11.1	<b>82.2</b>
AO1	$all(B, A); some-not(C, B)$	0.0	4.4	8.9	66.7	<b>17.8</b>	0.0	0.0	10.0	60.0	<b>30.0</b>	0.0	4.4	0.0	20.0	<b>75.6</b>
AO2	$all(A, B); some-not(C, B)$	0.0	4.4	4.4	<b>75.6</b>	15.6	0.0	3.3	6.6	<b>66.6</b>	23.3	0.0	0.0	2.2	<b>91.1</b>	6.7
AO3	$all(B, A); some-not(B, C)$	0.0	2.2	17.8	53.3	<b>26.7</b>	0.0	3.3	6.6	20.0	<b>70.0</b>	0.0	0.0	8.9	6.7	<b>77.8</b>
AO4	$all(A, B); some-not(B, C)$	0.0	4.4	11.1	55.6	<b>28.9</b>	0.0	3.3	0.0	13.3	<b>83.3</b>	0.0	2.2	0.0	6.7	<b>91.1</b>
OA1	$some-not(B, A); all(C, B)$	2.2	2.2	4.4	66.7	<b>24.4</b>	0.0	0.0	6.6	30.0	<b>63.3</b>	0.0	4.4	0.0	13.3	<b>82.2</b>
OA2	$some-not(A, B); all(C, B)$	2.2	2.2	4.4	64.4	<b>26.7</b>	0.0	6.6	3.3	26.6	<b>63.3</b>	0.0	6.7	2.2	8.9	<b>82.2</b>
OA3	$some-not(B, A); all(B, C)$	0.0	4.4	11.1	<b>80.0</b>	4.4	0.0	0.0	3.3	<b>60.0</b>	36.6	0.0	4.4	4.4	<b>66.7</b>	24.4
OA4	$some-not(A, B); all(B, C)$	0.0	11.1	20.0	42.2	<b>26.7</b>	3.3	3.3	6.6	16.6	<b>66.6</b>	0.0	2.2	4.4	2.2	<b>91.1</b>
EI1	$no(B, A); some(C, B)$	0.0	22.2	2.2	<b>62.2</b>	13.3	0.0	8.9	0.0	<b>84.4</b>	6.7	0.0	16.6	0.0	<b>70.0</b>	13.3
EI2	$no(A, B); some(C, B)$	0.0	26.7	4.4	<b>46.7</b>	22.2	3.3	10.0	0.0	<b>73.3</b>	10.0	0.0	8.9	0.0	<b>84.4</b>	6.7
EI3	$no(B, A); some(B, C)$	0.0	20.0	2.2	<b>53.3</b>	17.8	0.0	20.0	0.0	<b>63.3</b>	13.3	0.0	4.4	0.0	<b>84.4</b>	11.1
EI4	$no(A, B); some(B, C)$	0.0	28.9	6.7	<b>35.6</b>	28.9	0.0	10.0	0.0	<b>73.3</b>	16.6	0.0	6.7	0.0	<b>75.6</b>	15.6

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



Table B.2 Response distributions for 31 syllogisms in the Linear group, Symbolic group and 3-Venn group (Bold type refers to valid conclusion)

code $\mathcal{C}$ figure	premises 1st, 2nd	Linear group, N=21					Symbolic group, N=25					3-Venn group, N=38				
		conclusion*					conclusion*					conclusion*				
		A	E	I	O	N	A	E	I	O	N	A	E	I	O	N
AA2	$all(A, B); all(C, B)$	4.8	0.0	4.8	0.0	<b>90.5</b>	28.0	0.0	8.0	0.0	<b>64.0</b>	21.1	0.0	13.2	2.6	<b>63.2</b>
AA3	$all(B, A); all(B, C)$	4.8	0.0	33.3	0.0	<b>61.9</b>	28.0	0.0	44.0	0.0	<b>28.0</b>	18.4	0.0	10.5	0.0	<b>71.1</b>
AA4	$all(A, B); all(B, C)$	0.0	0.0	33.3	0.0	<b>66.7</b>	36.0	4.0	32.0	0.0	<b>28.0</b>	15.8	5.3	5.3	2.6	<b>71.1</b>
AI1	$all(B, A); some(C, B)$	0.0	0.0	<b>95.2</b>	0.0	4.8	0.0	0.0	<b>84.0</b>	4.0	12.0	0.0	0.0	<b>100.0</b>	0.0	0.0
AI2	$all(A, B); some(C, B)$	0.0	0.0	9.5	4.8	<b>85.7</b>	0.0	0.0	40.0	12.0	<b>48.0</b>	2.8	0.0	36.1	2.8	<b>58.3</b>
AI3	$all(B, A); some(B, C)$	0.0	0.0	<b>85.0</b>	0.0	15.0	4.0	0.0	<b>84.0</b>	0.0	12.0	2.6	0.0	<b>84.2</b>	2.6	10.5
AI4	$all(A, B); some(B, C)$	0.0	0.0	14.3	0.0	<b>85.7</b>	0.0	0.0	40.0	4.0	<b>56.0</b>	0.0	0.0	38.9	8.3	<b>52.8</b>
IA1	$some(B, A); all(C, B)$	0.0	0.0	19.0	0.0	<b>81.0</b>	4.0	4.0	44.0	0.0	<b>48.0</b>	2.6	0.0	55.3	0.0	<b>42.1</b>
IA2	$some(A, B); all(C, B)$	0.0	0.0	19.0	0.0	<b>81.0</b>	0.0	0.0	40.0	4.0	<b>56.0</b>	0.0	2.7	43.2	0.0	<b>54.1</b>
IA3	$some(B, A); all(B, C)$	0.0	0.0	<b>81.0</b>	0.0	19.0	0.0	0.0	<b>92.0</b>	0.0	8.0	0.0	0.0	<b>94.7</b>	0.0	5.
IA4	$some(A, B); all(B, C)$	0.0	0.0	<b>85.7</b>	0.0	14.3	0.0	0.0	<b>76.0</b>	4.0	20.0	0.0	0.0	<b>81.6</b>	2.6	13.2
AE1	$all(B, A); no(C, B)$	0.0	14.3	0.0	4.8	<b>81.0</b>	0.0	32.0	0.0	8.0	<b>60.0</b>	2.6	18.4	0.0	2.6	<b>76.3</b>
AE2	$all(A, B); no(C, B)$	0.0	<b>100.0</b>	0.0	0.0	0.0	0.0	<b>92.0</b>	0.0	0.0	8.0	0.0	<b>94.7</b>	5.3	0.0	0.0
AE3	$all(B, A); no(B, C)$	0.0	9.5	4.8	4.8	<b>81.0</b>	0.0	32.0	4.0	16.0	<b>48.0</b>	0.0	13.2	0.0	0.0	<b>86.7</b>
AE4	$all(A, B); no(B, C)$	0.0	<b>95.2</b>	0.0	4.8	0.0	0.0	<b>84.0</b>	0.0	4.0	12.0	0.0	<b>89.2</b>	0.0	5.4	5.4
EA1	$no(B, A); all(C, B)$	0.0	<b>90.5</b>	0.0	4.8	4.8	0.0	<b>91.7</b>	0.0	0.0	8.3	0.0	<b>97.4</b>	0.0	0.0	2.6
EA2	$no(A, B); all(C, B)$	0.0	<b>100.0</b>	0.0	0.0	0.0	0.0	<b>88.0</b>	0.0	0.0	12.0	0.0	<b>94.7</b>	0.0	0.0	5.3
EA3	$no(B, A); all(B, C)$	0.0	9.5	0.0	4.8	<b>85.7</b>	4.0	28.0	4.0	32.0	<b>32.0</b>	0.0	18.4	2.6	5.3	<b>73.7</b>
EA4	$no(A, B); all(B, C)$	0.0	14.3	0.0	4.8	<b>81.0</b>	4.0	28.0	4.0	32.0	<b>32.0</b>	2.6	10.5	0.0	7.9	<b>78.9</b>
AO1	$all(B, A); some-not(C, B)$	0.0	0.0	9.5	14.3	<b>76.2</b>	0.0	4.0	16.0	56.0	<b>24.0</b>	0.0	0.0	21.6	21.6	<b>56.7</b>
AO2	$all(A, B); some-not(C, B)$	0.0	0.0	0.0	<b>85.7</b>	14.3	0.0	8.3	0.0	<b>58.3</b>	33.3	0.0	0.0	2.7	<b>91.9</b>	5.4
AO3	$all(B, A); some-not(B, C)$	0.0	0.0	9.5	19.0	<b>71.4</b>	0.0	0.0	20.0	36.0	<b>44.0</b>	0.0	0.0	0.0	18.4	<b>81.6</b>
AO4	$all(A, B); some-not(B, C)$	0.0	4.8	0.0	14.3	<b>81.0</b>	0.0	4.0	4.0	44.0	<b>48.0</b>	0.0	2.7	0.0	18.9	<b>78.4</b>
OA1	$some-not(B, A); all(C, B)$	0.0	4.8	0.0	14.3	<b>81.0</b>	0.0	0.0	0.0	36.0	<b>64.0</b>	0.0	5.4	2.7	51.4	<b>40.5</b>
OA2	$some-not(A, B); all(C, B)$	0.0	4.8	0.0	9.5	<b>85.7</b>	0.0	8.0	4.0	44.0	<b>44.0</b>	0.0	2.6	0.0	31.5	<b>65.8</b>
OA3	$some-not(B, A); all(B, C)$	0.0	0.0	4.8	<b>47.6</b>	47.6	4.0	0.0	20.0	<b>64.0</b>	12.0	0.0	2.6	5.3	<b>78.9</b>	13.2
OA4	$some-not(A, B); all(B, C)$	0.0	0.0	9.5	9.5	<b>81.0</b>	0.0	0.0	12.0	44.0	<b>44.0</b>	0.0	0.0	26.3	15.8	<b>57.9</b>
EI1	$no(B, A); some(C, B)$	0.0	0.0	0.0	<b>85.7</b>	14.3	0.0	12.0	0.0	<b>48.0</b>	40.0	0.0	5.3	0.0	<b>92.1</b>	2.6
EI2	$no(A, B); some(C, B)$	0.0	9.5	0.0	<b>76.2</b>	14.3	0.0	4.0	0.0	<b>72.0</b>	24.0	0.0	7.9	0.0	<b>86.8</b>	5.3
EI3	$no(B, A); some(B, C)$	0.0	9.5	0.0	<b>66.7</b>	28.6	0.0	4.0	4.0	<b>64.0</b>	28.0	0.0	5.3	2.6	<b>76.3</b>	15.8
EI4	$no(A, B); some(B, C)$	0.0	4.8	0.0	<b>71.4</b>	23.8	0.0	4.0	0.0	<b>48.0</b>	44.0	0.0	10.0	0.0	<b>73.3</b>	16.6

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

# References

- Agnoli, F. (1991). Development of judgmental heuristics and logical reasoning: Training counteracts the representativeness heuristic. *Cognitive Development*, 6, 195–217.
- Allwein, G. & Barwise, J. (Eds.). (1996). *Logical Reasoning with Diagrams*. New York: Oxford University Press.
- Allwood, J., Anderson, L.-G., & Dahl, Ö. (1977). *Logic in Linguistics*. Cambridge, UK: Cambridge University Press.
- Bacon A., Handley S., & Newstead S. (2003). Individual differences in strategies for syllogistic reasoning. *Thinking and Reasoning*, 9, 133–168.
- Badre, D. & D’Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience*, 10, 659–669.
- 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. (1986/1993). *Turing’s World*, Santa Barbara, CA: Kinko’s Academic Courseware Exchange. Revised in (1993). Stanford, CA: CSLI Publications.
- Barwise, J. & Etchemendy, J. (1987/1991). *Tarski’s World*, Santa Barbara, CA: Kinko’s Academic Courseware Exchange. Revised in (1991). Stanford, CA: CSLI Publications.
- 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, No. 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.
- Barwise, J. & Shimojima, A. (1995). Surrogate reasoning. *Cognitive Studies: Bulletin of Japanese Cognitive Science Society*, 4, 7–27.
- Bauer, M. & Johnson-Laird, P.N. (1993). How diagrams can improve reasoning. *Psychological Science*, 4, 372–378.
- 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.
- Braine, M.D.S. & O’Brien, D.P. (Eds.) (1998). *Mental Logic*. Mahwah, NJ: Erlbaum.
- 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.
- Carroll, L. (1896). *Symbolic Logic*. New York: Dover.
- Chapman, L. & Chapman, J. (1959). Atmosphere effect re-examined. *Journal of Experimental Psychology*, 58, 220–226.
- Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, 38, 191–258.
- Chow, S. & Ruskey, F. (2004). Drawing area proportional Venn and Euler diagrams. In G. Liotta (Ed.), *Proceedings of Graph Drawing 2003, LNCS 2912* (pp. 466–477), Berlin Heidelberg: Springer.
- Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J.K., Holyoak, K.J., & Gabrieli, J.D.E. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage*, 14, 1136–1149.
- Copi, I.M. (1953). *Introduction to Logic*. New York: Macmillan.
- 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.
- Dummett, M. (1973). The justification of deduction. In M. Dummett (1978),

- Truth and Other Enigmas* (pp. 290–318), London: Duckworth.
- Englebrechtsen, 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-Petersbourg: 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.St.B.T., Barston, J., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition*, 11, 295–306.
- Evans, J.St.B.T., Newstead, S.E., & Byrne, R.J. (1993). *Human Reasoning: The Psychology of Deduction*. Hove: Erlbaum.
- Eysink, T.H.S., Dijkstra, S. & Kuper, J. (2001). Cognitive processes in solving variants of computer-based problems used in logic teaching. *Computers in Human Behavior*, 17, 1–19.
- 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.
- Gärdenfors, P. (2000). *Conceptual Spaces: On the Geometry of Thought*. Cambridge, MA: MIT Press.
- Gardner, M. (1958). *Logic Machines and Diagrams*. New York: McGraw-Hill.
- Gattis, M. (2004). Mapping relational structure in spatial reasoning. *Cognitive Science*, 28, 589–610.
- Geach, P.T. (1962). *Reference and Generality: An Examination of Some Medieval and Modern Theories*. Ithaca, NY: Cornell University Press.
- Geach, P.T. (1976). *Reason and Argument*. Oxford: Blackwell.
- George, C. (1997). Reasoning from uncertain premises. *Thinking and Reasoning*, 3, 161–189.
- Gergonne, J.D. (1817). Essai de dialectique rationnelle. *Annales de Mathématiques pures et appliquées*, 7, 189–228.
- Geurts, B. (2003a). Reasoning with quantifiers. *Cognition*, 86, 223–251.
- Geurts, B. (2003b). Monotonicity and syllogistic inference: A reply to Newstead. *Cognition*, 90, 201–204.

- Geurts, B. (2007). Existential import. In I. Comorovski & K. von Heusinger (Eds.), *Existence: Semantics and Syntax* (pp. 253–271). Dordrecht: Springer.
- Giaquinto, M. (1993). Diagrams: Socrates and Meno’s slave. *International Journal of Philosophical Studies*, 1, 81–97.
- Glasgow, J., Narayanan, N.H., & Chandrasekaran, B. (Eds.). (1995). *Diagrammatic Reasoning: Cognitive & Computational Perspectives*. Cambridge, MA: AAAI Press / MIT Press.
- Glymour, C. (1992). *Thinking Things Through: An Introduction to Philosophical Issues and Achievements*. Cambridge, MA: MIT Press.
- Goel, V. (2007). Anatomy of deductive reasoning. *Trends in Cognitive Sciences*, 11, 435–441.
- Goel, V., Buchel, C., Frith, C., & Dolan, R. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *NeuroImage*, 12, 504–514.
- Goel, V. & Dolan, R.J. (2003). Explaining modulation of reasoning by belief. *Cognition*, 87, B11–B22.
- Goel, V. & Dolan, R.J. (2004). Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition*, 93, 109–121.
- Goel, V., Tierney, M., Sheesley, L., Bartolo, A., Vartanian, O., & Grafman, J. (2007) Hemispheric specialization in human prefrontal cortex for resolving certain and uncertain inferences. *Cerebral Cortex*, 17, 2245–2250.
- Greaves, M. (2002). *The Philosophical Status of Diagrams*. Stanford, CA: CSLI Publications.
- Greenfield, P.M, Nelson, K. & Saltzman, E. (1972). The development of rule-bound strategies for manipulating seriated cups: A parallel between action and grammar. *Cognitive Psychology*, 3, 291–310.
- Gurr, C.A. (1998). On the isomorphism, or lack of it, of representations. In K. Marriot & B. Meyer (Eds.), *Theory of Visual Languages* (pp. 293–305), Berlin: Springer.
- Gurr, C.A. (1999). Effective diagrammatic communication: Syntactic, semantic and pragmatic issues. *Journal of Visual Languages & Computing*, 10, 317–342.
- Gurr, C.A., Lee, J., & Stenning, K. (1998). Theories of diagrammatic reasoning: Distinguishing component problems. *Minds and Machines*, 8, 533–557.
- Hackl, M. (2009). On the grammar and processing of proportional quantifiers: most versus more than half. *Natural Language Semantics*, 17, 63–98.
- 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. (1995). *Logic and Visual Information*. Stanford, CA: CSLI Publications.
- Hammer, E. & Shin, S. (1998). Euler’s visual logic. *History and Philosophy of Logic*, 19, 1–29.
- Hanakawa, T., Honda, M., Sawamoto, N., Okada, T., Yonekura, Y., Fukuyama, H., & Shibasaki, H. (2002). The role of rostral Brodmann area 6 in mental operation tasks: An integrative neuroimaging approach. *Cerebral Cortex*, 12, 1157–1170.
- Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1084–1102.
- 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.
- Howse, J., Stapleton, G., & Taylor, J. (2005). Spider diagrams. *LMS Journal of Computation and Mathematics*, 8, 145–194.
- Johnson-Laird, P.N. (1983). *Mental Models : Towards Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press.
- Johnson-Laird, P.N., & Bara, B. (1984). Syllogistic inference. *Cognition*, 16, 1–61.
- Johnson-Laird, P.N., & Byrne, R. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.
- Johnson-Pynn, J., Frigaszy, 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.
- Jonides, J., Schumacher, E.H., Smith, E., Lauber, E.J., Awh, E., Minoshima, S., & Koeppe, R. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience*, 9, 462–475.
- Karnaugh, M. (1953). The map method for synthesis of combinational logic circuits. *Transactions of the American Institute of Electrical Engineers, part I*, 72, 593–599.
- 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.
- Khemlani, S., Leslie, S.J., & Glucksberg, S. (2012). Inferences about members of kinds: The generics hypothesis. *Language and Cognitive Processes*, 27, 887–900.
- Knauff, M. (2007). How our brains reason logically. *Topoi*, 26, 19–36.
- Kosslyn, S.M. (1994). *Elements of Graph Design*. New York: W.H. Freeman and Company.
- Lakoff, G. & Núñez, R.E. (2000). *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. New York: Basic Books.
- 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.
- Larkin, J. & Simon, H. (1987). Why a diagram is (sometimes) worth 10,000 words. *Cognitive Science*, 11, 65–99.
- Lear, J. (1986). *Aristotle and Logical Theory*. Cambridge, England: Cambridge University Press.
- Leibniz, G.W. (1903/1988). *Opusculs et fragments inédits: 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.
- Leslie, S. J., Khemlani, S., & Glucksberg, S. (2011). Do all ducks lay eggs? The generic overgeneralization effect. *Journal of Memory and Language*, 65, 15–31.
- Levesque, H.J. (1988). Logic and the complexity of reasoning. *Journal of Philosophical Logic*, 17, 335–389.
- Lowe, R. & Schnotz, W. (Eds.). (2008). *Learning with Animation: Research Implications for Design*. New York: Cambridge University Press.
- Lull, R. (1617). *Opera Ea Quae Ad Adinventam Ab Ipso Artem Universalem, Scientiarum Artiumque Omnium Breui Compendio, Firmaq; Memoria Apprehendendarum, Locupletiffimaq; Vel Oratione Ex Tempore Pertractandarum, Pertinent*. Strassburg: Lazari Zetzneri.
- McCarthy, J. (1980). Circumscription—a form of non-monotonic reasoning, *Artificial Intelligence*, 13, 27–39.

- McCarthy, J. (1986). Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence*, 28, 89–116.
- Mercier, H. & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Science*, 34, 57–111.
- 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. (2010). Two types of diagrammatic inference systems: Natural deduction style and resolution style. In A. Goel, M. Jamnik, & N.H. Narayanan (Eds.), *Proceedings of Diagrams 2010, LNAI 6170*, (pp. 99–114), 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. (2012). On the cognitive efficacy of Euler diagrams in syllogistic reasoning: a relational perspective. In P. Chapman, & L. Micallef (Eds.), *Proceedings of Euler Diagrams 2012, CEUR Workshop Proceedings 854* (pp. 17–31).
- 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.
- Montague, R. (1974). *Formal Philosophy: Selected Papers of Richard Montague*, In R.H. Thomason (Ed.). New Haven, CT: Yale University Press.
- Monti, M.M., Osherson, D.N., Martinez, M.J., & Parsons, L.M. (2007). Functional neuroanatomy of deductive inference: A language-independent distributed network. *NeuroImage*, 37, 1005–1016.
- Monti, M.M., Parsons, L.M., & Osherson, D.N. (2009). The boundaries of language and thought in deductive inference. *Proceedings of the National Academy of Sciences of the U.S.A.*, 106, 12554–12559.
- Morgan, J.J.B., & Carrington, D.H. (1944). Graphic instruction in relational reasoning. *Journal of Educational Psychology*. 35, 536–544.



- Moss, L.S. (2010). *Natural logic and semantics*. In M. Aloni, H. Bastiaanse, T. Jäger, K. Schulz (Eds.), *Proceedings of Amsterdam Colloquium 2009, LNAI 6042* (pp. 84–93). Berlin Heidelberg: Springer.
- Narayanan, N. (Ed.). (1992). *Working Notes of AAAI Spring Symposium on Reasoning with Diagrammatic Representations*, Menlo Park, CA: AAAI Press.
- Newstead, S.E. & Griggs, R. (1983). Drawing inferences from quantified statements: A study of the square of opposition. *Journal Verbal Learning and Verbal Behavior*, 22, 535–546.
- Norman, D.A. (1993). *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Reading, MA: Addison-Wesley Publishing.
- Noveck, I.A., Goel, V. & Smith, K.W. (2004). The neural basis of conditional reasoning with arbitrary content. *Cortex*, 40, 613–622.
- Novick, L.R. (2006). Understanding spatial diagram structure: An analysis of hierarchies, matrices, and networks. *Quarterly Journal of Experimental Psychology*, 59, 1826–1856.
- Novick, L.R. & Hurley, S.M. (2001). To matrix, network, or hierarchy: That is the question. *Cognitive Psychology*, 42, 158–216.
- Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, 5, 349–357.
- Oberauer, K., & Wilhelm, O. (2000). Effects of directionality in deductive reasoning: I. The comprehension of single relational premises. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1702–1712.
- Oberauer, K., Hörnig, R., Weidenfeld, A., & Wilhelm, O. (2005). Effects of directionality in deductive reasoning: II. Premise integration and conclusion evaluation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 58A, 1225–1247.
- 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.
- Pietroski, P., Lidz, J., Hunter, T., & Halberda, J. (2009). The meaning of ‘most’: Semantics, numerosity and psychology. *Mind and Language*, 24, 554–585.
- Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., & Gabrieli J.D.E. (1999). Functional specialization for semantic and phonological processing in the left inferior frontal cortex. *NeuroImage*, 10, 15–35.
- Politzer, G. (2011). Solving natural syllogisms. In K. Manktelow, D. Over, & S. Elqayam (Eds.), *The Science of Reason: A Festschrift for Jonathan St. BT Evans* (pp. 19–35), New York: Psychology Press.
- Politzer, G. & Mercier, H. (2008). Solving categorical syllogisms with singular

- premises. *Thinking and Reasoning*, 14, 414–453.
- 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.
- Prado, J., Chadha, A., & Booth, J.R. (2011). The brain network for deductive reasoning: A quantitative meta-analysis of 28 neuroimaging studies. *Journal of Cognitive Neuroscience*, 23, 3483–3497.
- Ramnani, N. & Owen, A.M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience*, 5, 184–194.
- Ratwani, R.M., Trafton, J.G., & Boehm-Davis, D.A. (2008). Thinking graphically: Connecting vision and cognition during graph comprehension. *Journal of Experimental Psychology: Applied*, 14, 36–49.
- Rescher, N. & Gallagher, N. (1965). Venn diagrams for plurative syllogisms. *Philosophical Studies*, 16, 49–55.
- Reverberi, C., Cherubini, P., Frackowiak, R., Catagirone, C., Paulesu, E., & Macaluso, E. (2010). Conditional and syllogistic deductive tasks dissociate functionally during premise integration. *Human Brain Mapping*, 31, 1430–1445.
- Reverberi, C., Bonatti, L., Frackowiak, R., Paulesu, E., Cherubini, P., & Macaluso, E. (2012). Large scale brain activation predict reasoning profile. *NeuroImage*, 59, 1752–1764.
- Revlis, R. (1975). Two models of syllogistic reasoning: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior*, 14, 180–195.
- Rips, L. (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking*. Cambridge, MA: MIT Press.
- 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. & Sykes, E.D.A. (2005). Categorical reasoning from multiple diagrams. *The Quarterly Journal of Experimental Psychology*, 58A(2), 333–376.
- Rodriguez-Morenoa, D. & Hirschb, J. (2009). The dynamics of deductive reasoning: An fMRI investigation. *Neuropsychologia*, 47, 949–961.
- Sato, Y., Takemura, R., Mineshima, K., Shikishima, C., Sugimoto, Y., Ando, J., & Okada, M. (2008). Some remarks on deductive syllogistic reasoning studies: towards an integrated methodology of logical and psychological approaches. In M. Okada, R. Takemura & J. Ando (Eds.), *Reports on Interdisciplinary Logic Inference Studies* (pp. 3–32), Keio University Press.

- Sato, Y., Mineshima, K., & Takemura, R. (2010a). The efficacy of Euler and Venn diagrams in deductive reasoning: Empirical findings. In A. Goel, M. Jamnik, & N.H. Narayanan (Eds.), *Proceedings of Diagrams 2010, LNAI 6170*, (pp. 6–22), Berlin Heidelberg: Springer.
- Sato, Y., Mineshima, K., & Takemura, R. (2010b). 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. (2010c). Children’s reasoning with logic diagrams. In David N. Rapp (Ed.), *Spatial Cognition 2010: Poster Presentations* (pp. 81–84). SFB/TR 8 Report, No. 024–07/2010.
- 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., & Mineshima, K. (2012). The efficacy of diagrams in syllogistic reasoning: A case of linear diagrams. In P. Rodgers, P. Cox, & B. Plimmer (Eds.), *Proceedings of Diagrams 2012, LNAI 7352* (pp. 352–355), Berlin Heidelberg: Springer.
- Sato, Y., Masuda, S., Someya, Y., Tsujii, T., & Watanabe, S. (in preparation). An fMRI analysis of the efficacy of external diagrams in deductive reasoning: Changes in brain activation of the prefrontal cortex.
- Scaife, M. & Rogers, Y. (1996). External cognition: how do graphical representations work?. *International Journal of Human-Computer Studies*, 45, 185–213.
- Schwartz, S. (1971). Modes of representation and problem solving: Well evolved is half solved. *Journal of Experimental Psychology*, 91, 347–350.
- Schwartz, S. & Fattaleh, D. (1972). Representation in deductive problem solving: The matrix. *Journal of Experimental Psychology*, 95, 343–348.
- Schwartz, D.L. (1995). Reasoning about the referent of a picture versus reasoning about the picture as the referent: An effect of visual realism. *Memory and Cognition*, 23, 709–722.
- Shimojima, A. (1996). *On the Efficacy of Representation*. PhD thesis, Indiana University.
- Shimojima, A. (1999a). Constraint-preserving representations. In L. Moss, J. Ginzburg, & M. de Rijke, (Eds), *Logic, Language and Computation: Volume 2, CSLI Lecture Notes No.96* (pp. 296–317). Stanford, CA: CSLI Publications.
- Shimojima, A. (1999b). The graphic linguistic distinction. *Artificial Intelligence Review*, 13, 313–35.

- Shimojima, A. & Fukaya, T. (2003). Do we really reason about a picture the referent. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 1076–1081), Austin, TX: Cognitive Science Society.
- Shimojima, A. & Katagiri, Y. (2008a). Hypothetical drawing in embodied spatial reasoning. In B.C. Love, K. McRae, & V.M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2247–2252), Austin, TX: Cognitive Science Society.
- Shimojima, A. & Katagiri, Y. (2008b). An eye-tracking study of exploitations of spatial constraints in diagrammatic reasoning. In J. Howse, & J. Lee (Eds.), *Proceedings of Diagrams 2008, LNAI 5223* (pp. 74–88), Berlin Heidelberg: Springer.
- Shimojima, A. & Katagiri, Y. (2010). An eye-tracking study of integrative spatial cognition over diagrammatic representations. In C. Hölscher et al. (Eds.), *Spatial Cognition VII, LNAI 6222* (pp. 262–278), Berlin Heidelberg: Springer.
- Shin, S.-J. (1994). *The Logical Status of Diagrams*. New York: Cambridge University Press.
- Shin, S.-J. & Lemon, O. (2001/2008). Diagrams. *Stanford Encyclopedia of Philosophy*, Available at: <http://plato.stanford.edu/entries/diagrams/>
- Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., Evans, A., Rapaport, J., & Giedd, J. (2006). Intellectual ability and cortical development in children and adolescent. *Nature*, 440, 676–679.
- Sloman, S.A., Over, D., Slovak, L., & Stibel, J.M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes*, 91, 296–309.
- Sloutsky, V.M. & Lo, Y-F. (1999). How much does a shared name make things similar? Part 1: Linguistic labels and the development of similarity judgment. *Developmental Psychology*. 35, 1478–1492.
- Smith, R. (1982). What is Aristotelian Ecthesis? *History and Philosophy of Logic*, 3, 113–127.
- Stanovich, K.E. & West, R.F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, 23, 645–726.
- Stapleton, G. (2005). A survey of reasoning systems based on Euler diagrams. *Proceedings of Euler Diagrams 2004, ENTCS 134* (pp. 127–151). Amsterdam: Elsevier.
- Stapleton, G., Rodgers, P., & Howse, J. (2011). A general method for drawing area-proportional Euler diagrams. *Journal of Visual Languages and Computing*, 22, 426–442.

- Stenning, K. (2000). Distinctions with differences: Comparing criteria for distinguishing diagrammatic from sentential systems. In M. Anderson, P. Cheng, & Haarslev (Eds.), *Proceedings of Diagrams 2000, LNAI 1889* (pp. 132–148), Berlin Heidelberg: Springer.
- Stenning, K. (2002). *Seeing Reason: Image and Language in Learning to Think*. Oxford: Oxford University Press.
- Stenning, K. & Cox, R. (2006). Reconnecting interpretation to reasoning through individual differences. *The Quarterly Journal of Experimental Psychology*, 59, 1454–1483.
- 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., & Yule, P. (1997). Image and language in human reasoning: Syllogistic illustration. *Cognitive Psychology*, 34, 109–159.
- Stenning, K. & van Lambalgen, M. (2008). *Human Reasoning and Cognitive Science*. Cambridge, MA: MIT Press.
- Stevenson, R. J., & Over, D. E. (1995). Deduction from uncertain premises. *The Quarterly Journal of Experimental Psychology*, 48A, 613–643.
- Störring, G. (1908). Experimentelle untersuchungen über einfache schlussprozesse. *Archiv für die gesamte Psychologie*, 11, 1–127.
- Strawson, P.F. (1952). *Introduction to Logical Theory*. London: Methuen.
- Szymanik, J. & Zajenkowski, M. (2010). Comprehension of simple quantifiers: Empirical evaluation of a computational model. *Cognitive Science*, 34, 521–532.
- Tennant, N. (1986). The withering away of formal semantics? *Mind and Language*, 1, 302–318.
- Trafton, J.G., Trickett, S.B., & Mintz, F.E. (2005). Connecting internal and external representations: Spatial transformations of scientific visualizations. *Foundations of Science*, 10, 89–106.
- Tsujii, T., Sakatani, K., Masuda, S., Akiyama, T., & Watanabe, S. (2011). Evaluating the roles of the inferior frontal gyrus and superior parietal lobule in deductive reasoning: an rTMS study, *NeuroImage*, 58, 640–646.
- van Benthem, J. (2008). *A Brief History of Natural Logic, Technical Report PP-2008-05*, Institute for Logic, Language and Computation. University of Amsterdam.
- van den Heuvel, O.A., Groenewegen, H.J., Barkhof, F., Lazeron, R.H.C., van

- Dyck, R., & Veltman, D.J. (2003). Frontostriatal system in planning complexity: a parametric functional magnetic resonance version of Tower of London task. *NeuroImage*, 18, 367–374.
- Venn, J. (1881). *Symbolic Logic*. London: Macmillan.
- Wickes, W.E. (1968). *Logic Design with Integrated Circuits*. New York: John Wiley & Sons.
- Yamagishi, K. (2003). Facilitating normative judgments of conditional probability: Frequency or nested sets? *Experimental Psychology*, 50, 97–106.
- Zhang, J., & Norman, D.A. (1994). Representations in distributed cognitive tasks. *Cognitive Science*, 18, 87–122.
- Zielinski, T., Goodwin, G., & Halford, G. (2010). Complexity of categorical syllogisms: An integration of two metrics. *European Journal of Cognitive Psychology*, 22, 391–421.