Evaluations of Epistemic Components for Resolving the Muddy Children Puzzle*

J. Jude Kline †

19 November 2012

Abstract

We evaluate the 3 child muddy children puzzle using the epistemic logic of shallow depths GL_{EF}. This system is used to evaluate what components are necessary for a resolution. These components include the basic beliefs of a child, the necessary depths of the epistemic structures, and the observations about the inactions of others added after a stage. These are all given explicitly, and their necessity is examined. We formulate the concept of a resolution as a process of inferences, actions, observations, and belief changes. We give three main theorems. The first one gives a specific resolution, in which no common knowledge is involved. The second theorem states that any resolution has length of at least 3. The third theorem shows that the resolution given in the first theorem is minimal in various senses. In this manner, the necessary components for a resolution of the puzzle are evaluated. A final theorem gives a resolution for the n-child case.

JEL Classification: D80, D83

Keywords: Syntactic Approach, Shallow Depths of Epistemic Structures, Gentzen-style Sequent Calculus, Classical and Epistemic Inferences

1. Introduction

The muddy children puzzle is well known in game theory and logic. It appears in the literature in various forms, e.g., the cheating husbands puzzle, or colored hats puzzle, but with, more or less, the same logical structure. The general situation of the muddy children puzzle is that the children in a class are all muddy, and each can see everyone else's face, but not his own. The teacher announces that at least one child has a muddy

*I am very grateful to Mamoru Kaneko for developing this paper. I also thank Chi Chang, Amitaba Mukerjee, and Kisaragi for comments and encouragement. Finally, I thank Andrei Trandafira for motivating my interest in pursuing this problem. Finally, I thank two anonymous referees for suggestions and questions that helped to improve the paper.

†School of Economics, University of Queensland, Brisbane, QLD 4072, Australia, (j.kline@uq.edu.au)

face, and he asks each child to raise his hand if he concludes that his face is muddy. At this point of time, no child raises his hand. After some time observing the inactions of others, however, each child raises his hand. This is regarded as a puzzle in that the teacher's announcement together with observations and logical inferences leads to an unexpected result.

We find various logical treatments of the puzzle: Bollobas [4] gave an early description of the puzzle in English; Barwise [2], Gerbrandy-Groeneveld [8] and van Ditmarsch, *et al.* [18] among others gave logical (semantical) treatments of the puzzle; Geanakoplos-Polemarchakis [5], Binmore [3], and Geanakoplos [6] among others gave an information partitional approach to it. Those papers look for some formulations of how the children conclude their faces are muddy. However, the puzzle involves quite subtle, intra/interpersonal, logical inferences. In fact, it contains, in addition to pure logical inferences, observations of actions and their interpretations. Those aspects are often implicit and entangled in the above-mentioned approaches. In this paper, we try to separate those aspects taking the 3 child case.

We now give a small summary of the main results of this paper. After that, we describe some limitations of the above-mentioned approaches. Finally, we describe some important differences in our approach. We give three theorems in Section 3. The first one gives a specific resolution, in which only the depth 3 of interpersonal beliefs are required, *a fortiori*, no common knowledge is involved. The second theorem states that any resolution has length of at least 3. The third theorem shows that the resolution given in the first theorem is minimal in various senses. In this manner, the necessary components for a resolution of the puzzle are evaluated. The point of the paper is not simply to show this resolution, but to evaluate and find what components are necessary for a resolution.

In order to emphasize the points of this paper, now we discuss some limitations of the information partition approach, and also the semantical approaches.

(1) **Logical inferences are implicit:** In the information partition approach to the puzzle, the main logical arguments remain informal; the inferences are all in the interpretations. Also, these interpretations mix intrapersonal and interpersonal logical inferences. While the semantical approaches give more precise treatments of these components, the inferences are still indirectly described by semantic models, and, specifically, the classical and epistemic inferences are not well separated. For further analysis and understanding, it would be important to separate these components.

(2) **What is the role of common knowledge?** In the information partition/semantical approaches, the announcement by the teacher as well as the observations of actions are formulated to be common knowledge. Moreover, the information partition approach relies upon a more basic implicit assumption that the information partitions themselves are common knowledge.

In the following, we describe two aspects of our approach in response to the issues raised in (1) and (2), and in addition, we give a comment on the treatment of updating basic beliefs.

The first aspect is our response to (1). We treat intra/interpersonal logical inferences explicitly, which are described syntactically in the epistemic logic GL_{EF} of Kaneko-Suzuki [14]. The subscripts E and F are constraints on the interpersonal belief hierarchies. The player at the innermost level of a hierarchy of beliefs is allowed to make classical logic inferences, and those inferences are considered by the players further out in the hierarchy.

For example, in Figure 1.1, child i makes classical logic (CL) inferences referring to j's CL inferences, and indirectly, through j, to k's CL inferences. Within GL_{EF}, we can analyze the precise classical inferences and belief hierarchies involved in the puzzle[1].

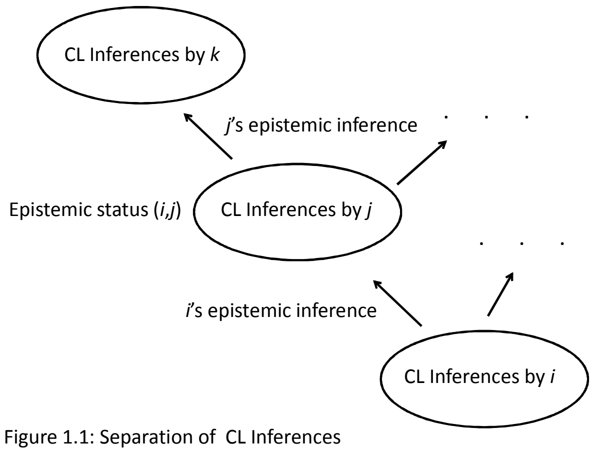


Figure 1.1: Separation of CL Inferences

The second aspect to be emphasized is about (2). As mentioned above, our approach requires only interpersonal beliefs up to depth 3, *a fortiori*, common knowledge is unnecessary. Sato [16] showed this result, but we evaluate the necessity of depth 3.[2] The logic GL_{EF} facilitates these types of evaluations by considering explicit bounds on the epistemic depths of inferences.

Now we give one comment on our treatment of updating. Before the puzzle, each child has a finite set of basic beliefs about the observational abilities of the other children,

[1]Suzuki [17] gave an intuitionistic version of GL_{EF} called IG_{EF}. The analysis of this paper may be done in IG_{EF}, though we may need to include more basic beliefs in order to obtain a resolution in IG_{EF}.

[2]Yasugi-Oda [19] gave a syntactic argument for the two children puzzle. However, their puzzle differs in that one child sees the other, but the other does not see the first.

the announcement by the teacher, and the action rule: "raise your hand if you reach the belief that your face is muddy". After the teacher's question, each child observes the actions (inactions) of others. Those observations may be adopted as new beliefs.

In the dynamic epistemic logic approach (cf. Gerbrandy-Groeneveld [8] and van Ditmarsch, *et al.* [18]), the process of updating is included in one formal system. However, we separate the process of updating from the logical system. We treat the entire situation as a time series of evolving basic beliefs, where a child considers his situation at each point of time taking his current basic beliefs as given and making logical inferences from them. Since, in this sense, logical inferences and the updating of basic beliefs have different time structures, we adopt our separate treatment.

To keep the discourse of the paper as clear as possible, we have restricted ourselves to the 3 child case. The 3 child case has a new feature which does not appear in the 2 child case: the accumulation of beliefs about observations of inactions appears in the 3 child case, but it is trivial in the 2 child case. Thus, we take the 3 child case. The main point of the paper is not only to give a specific resolution but also to explore the type of reasoning, evolution of beliefs, and epistemic inferences involved. We do this completely for the 3 child case. A generalization to the n child case is possible, but it makes the discourse too complicated to show our main points. Nevertheless, we do give a theorem about a resolution for the the n-person case in Section 4.

The organization of the paper is as follows. Section 2 describes the epistemic logic GL_{EF} that will be used to analyze the puzzle. In Section 3 we describe the general process of the 3 child puzzle and define a resolution to the puzzle. We give a resolution in Theorem 3.1, which makes the beliefs, inferences, epistemic structures, and interpretations explicit. We also give Theorems 3.2 and 3.3, which describe why this resolution may be regarded as a minimal one. Section 4 treats the n child case. We show, in Theorem 4.1, that the resolution given in Theorem 3.1 can be extended to the n child case. In Section 5 we discuss the results and give some concluding remarks. Section 6 gives two meta-theorems for epistemic logic that are also used to prove the theorems of Section 3. The proofs of Theorems 3.2 and 3.3 are given in Section 7.

2. Epistemic Logic of Shallow Depths

As mentioned in Section 1, one point of this paper is to show that the muddy children puzzle can be handled with shallow epistemic reasoning. Here, we describe the epistemic logic of shallow depths, following Kaneko-Suzuki [14]. In Section 2.1, we give the formal language used to describe the puzzle and the reasoning of the children. In Section 2.2, we define epistemic depths. In Section 2.3, we give a Gentzen-style formulation of epistemic logic GL_{EF}.

2.1. Language for the Muddy Children Puzzle

The language starts with the primitive propositional symbols $m_1, m_2, m_3, r_1, r_2, r_3$. The intended meaning of m_i is "player i's face is muddy", and the intended meaning of r_i is that "player i raises his hand". The primitive connective symbols are: $\neg$ (not), $\wedge$ (and), $\vee$ (or), and $\supset$ (implies). We also introduce the belief operators B_1, B_2, B_3 and use parentheses "(", ")", "{", "}" and comma ",". The formulas are inductively formed by:

(1.1) All primitive propositional symbols are formulas;

(1.2) If A and B are formulas, then so are $(\neg A)$, $(A \supset B)$, and $\mathrm{B}_i(A)$ for $i = 1, 2, 3$;

(1.3) If $\{A_1, ..., A_m\}$ is a finite nonempty set of formulas, then $\wedge\{A_1, ..., A_m\}$ and $\vee\{A_1, ..., A_m\}$ are formulas.

We denote the set of all formulas by $\mathcal{P}$. We eliminate parentheses when it does not cause any confusion, e.g., $(A \supset B)$ may be abbreviated by $A \supset B$. Also, we will may abbreviate $\wedge\{A, B\}$, $\wedge\{A, B, C\}, \vee\{A, B\}$, etc., as $A \wedge B$, $A \wedge B \wedge C$, and $A \vee B$.

We let $N^{<\omega} = \{(i_1, ..., i_m) : i_1, ..., i_m \in \{1, 2, 3\}\}$. We stipulate that $N^{<\omega}$ contains the *null sequence* ϵ: when $m = 0$, $(i_1, ..., i_m)$ is understood as ϵ. A sequence in $N^{<\omega}$ is called an *epistemic status.* For any $e = (i_1, ..., i_m) \in N^{<\omega}$, and any formula A in $\mathcal{P}$, we write $\mathrm{B}_e(A)$ for $\mathrm{B}_{i_1}...\mathrm{B}_{i_m}(A)$; when $e = \epsilon$, $\mathrm{B}_e(A) = A$. We also define the concatenation of $e = (i_1, ..., i_m)$ and $e' = (j_1, ..., j_k)$ in $N^{<\omega}$ by $e \circ e' = (i_1, ..., i_m, j_1, ..., j_k)$. We stipulate that $e \circ \epsilon = \epsilon \circ e = e$, and we write $(i) \circ e$ and $e \circ (i)$ as $i \circ e$ and $e \circ i$.

Now we define the *(epistemic) depth* $\delta(\cdot)$ inductively by:

D0: $\delta(A) = \{\epsilon\}$ if A is a primitive propositional symbol;

D1: $\delta(\neg A) = \delta(A)$;

D2: $\delta(A \supset B) = \delta(A) \cup \delta(B)$;

D3: $\delta(\wedge\Phi) = \delta(\vee\Phi) = \cup_{C\in\Phi}\delta(C)$;

D4: $\delta(\mathrm{B}_i(A)) = \{i \circ e : e \in \delta(A)\}$.

For example, the formula $\mathrm{B}_2(m_1) \supset m_1$ has epistemic depth $\delta(\mathrm{B}_2(m_1) \supset m_1) = \delta(\mathrm{B}_2(m_1)) \cup \delta(m_1) = \{(2), \epsilon\}$. For player 1's belief of this formula, i.e., $\mathrm{B}_1(\mathrm{B}_2(m_1) \supset m_1)$, the epistemic depth is $\delta(\mathrm{B}_1(\mathrm{B}_2(m_1) \supset m_1)) = \{(1, 2), (1)\}$. For a set of formulas Γ, we define $\delta(\Gamma) = \bigcup_{C\in\Gamma} \delta(C)$.

A non-empty subset E of $N^{<\omega}$ is an *epistemic structure* iff

$$(i_1, ..., i_m) \in E \text{ implies } (i_1, ..., i_{m-1}) \in E. \tag{2.1}$$

The null sequence ϵ is always in E by (2.1), since $(i_1, ..., i_{m-1}) = \epsilon$ for $m = 1$. Given an epistemic structure E, we define the set of *admissible formulas in* E as

$$\mathcal{P}_E = \{A \in \mathcal{P} : \delta(A) \subseteq E\}. \tag{2.2}$$

By the epistemic structure E, we restrict the set of all formulas $\mathcal{P}$ to the formulas having depths included in E. For the epistemic logic GL_{EF}, we will use another epistemic structure F, which is a subset of E, to control interpersonal inferences.

2.2. Epistemic Logic $\mathbf{GL}_{EF}$

The logical inferences of the players are described by a Gentzen-style formulation of the epistemic logic GL_{EF}. Here, the subscripts E and F are two epistemic structures. As in (2.2), E restricts the depths of beliefs in the language, and F restricts the depths of interpersonal inferences. We require $F \subseteq E$.

To describe interpersonal inferences, Kaneko-Suzuki [14] introduced the concept of a thought sequent. Let $e = (i_1, ..., i_m) \in E$, and let Γ, Θ denote finite (possibly empty) subsets of $\mathcal{P}_E$. Let [,] and $\rightarrow$ be auxiliary symbols used to form the new expression $\mathrm{B}_e[\Gamma \rightarrow \Theta]$, which we call a *thought sequent*. The expression $\mathrm{B}_e[\Gamma \rightarrow \Theta]$ represents the thought of the outer-most player i_1 on player i_2's thinking about...the inner most player i_m's logical reasoning.

We say that a thought sequent $\mathrm{B}_e[\Gamma \rightarrow \Theta]$ is *admissible* in E iff

$$e \circ \delta(\Gamma \cup \Theta) := \{e \circ e' : e' \in \delta(\Gamma \cup \Theta)\} \subseteq E. \tag{2.3}$$

This means that a thought sequent $\mathrm{B}_e[\Gamma \rightarrow \Theta]$ as a whole is allowed by E. For an epistemic status e, we define the set $\mathcal{P}_E(e) = \{A \in \mathcal{P} : e \circ \delta(A) \subseteq E\}$. A thought sequent $\mathrm{B}_e[\Gamma \rightarrow \Theta]$ is admissible in E if and only if $A \in \mathcal{P}_E(e)$ for all $A \in \Gamma \cup \Theta$. When $e = \epsilon$, $\mathrm{B}_\epsilon[\Gamma \rightarrow \Theta]$ is interpreted as the outside investigator's thought about the provability of $\Gamma \rightarrow \Theta$. Let Γ, Θ, Δ, and Λ be finite sets of formulas and A and C be formulas. We will abbreviate $\mathrm{B}_\epsilon[\Gamma \rightarrow \Theta]$ by $\Gamma \rightarrow \Theta$, $\mathrm{B}_e[\Gamma \cup \Delta \rightarrow \Theta \cup \Lambda]$ by $\mathrm{B}_e[\Gamma, \Delta \rightarrow \Theta, \Lambda]$, and $\mathrm{B}_e[\{A\} \cup \Gamma \rightarrow \Theta \cup \{C\}]$ by $\mathrm{B}_e[A, \Gamma \rightarrow \Theta, C]$, etc.

When Γ is empty in $\mathrm{B}_e[\Gamma \rightarrow \Theta]$, it is intended to mean that Θ follows logically without any assumption; and when Θ is empty, Γ leads to a contradiction.

Epistemic logic GL_{EF} is formally defined by one axiom schema and various inference rules. The inference rules are divided into three classes: structural rules, operational rules, and one epistemic rule. The structural and operational rules with the axiom schema are based on classical logic; in Figure 1.1, each oval represents the classical logic part within a fixed epistemic status, e.g., classical logic in (i, j) is represented by the middle oval. The epistemic rule connects those ovals, e.g., the top oval is connected to the middle oval.

Let $\Gamma, \Theta, \Delta, \Lambda, \Phi$ be arbitrary finite subsets of $\mathcal{P}_E(e)$ and $e \in F$, where Φ is non-empty, and let A and B be arbitrary formulas in $\mathcal{P}_E$:

Axiom (Initial Sequent): $\mathrm{B}_e[A \rightarrow A]$

Structural Rules:

$$\frac{\mathrm{B}_e[\Gamma \to \Theta]}{\mathrm{B}_e[\Delta, \Gamma \to \Theta, \Lambda]} \text{ (th)}$$

$$\frac{\mathrm{B}_e[\Gamma \to \Theta, A] \quad \mathrm{B}_e[A, \Delta \to \Lambda]}{\mathrm{B}_e[\Gamma, \Delta \to \Theta, \Lambda]} \text{ (cut)}$$

Operational Rules:

$$\frac{\mathrm{B}_e[\Gamma \to \Theta, A]}{\mathrm{B}_e[\neg A, \Gamma \to \Theta]} (\neg \to) \qquad \frac{\mathrm{B}_e[A, \Gamma \to \Theta]}{\mathrm{B}_e[\Gamma \to \Theta, \neg A]} (\to \neg)$$

$$\frac{\mathrm{B}_e[\Gamma \to \Theta, A] \quad \mathrm{B}_e[B, \Delta \to \Lambda]}{\mathrm{B}_e[A \supset B, \Gamma, \Delta \to \Theta, \Lambda]} (\supset\to) \qquad \frac{\mathrm{B}_e[A, \Gamma \to B, \Theta]}{\mathrm{B}_e[\Gamma \to A \supset B, \Theta]} (\to\supset)$$

$$\frac{\mathrm{B}_e[A, \Gamma \to \Theta]}{\mathrm{B}_e[\wedge\Phi, \Gamma \to \Theta]} (\wedge \to), \text{ where } A \in \Phi \qquad \frac{\{\mathrm{B}_e[\Gamma \to \Theta, A] : A \in \Phi\}}{\mathrm{B}_e[\Gamma \to \Theta, \wedge\Phi]} (\to \wedge)$$

$$\frac{\{\mathrm{B}_e[A, \Gamma \to \Theta] : A \in \Phi\}}{\mathrm{B}_e[\vee\Phi, \Gamma \to \Theta]} (\vee \to) \qquad \frac{\mathrm{B}_e[\Gamma \to \Theta, A]}{\mathrm{B}_e[\Gamma \to \Theta, \vee\Phi]} (\to \vee), \text{ where } A \in \Phi$$

Epistemic Distribution Rule: for $e \circ i \in F$,

$$\frac{\mathrm{B}_{e \circ i}[\Gamma \to \Theta]}{\mathrm{B}_e[\mathrm{B}_i(\Gamma) \to \mathrm{B}_i(\Theta)]} (\mathrm{B}_i \to \mathrm{B}_i), \text{where } |\Theta| \le 1, \text{and } i = 1, 2, 3$$

Here $|\Theta|$ denotes the cardinality of Θ.

A rule of inference, say $(\neg \to)$, means that if the upper sequent $\mathrm{B}_e[A, \Gamma \to \Theta]$ is proved, then the lower sequent $\mathrm{B}_e[\Gamma \to \Theta, \neg A]$ is inferred. A proof is a tree of thought sequents, each of which is connected by a rule of inference, having only instances of the axiom schema as its leaves. This basic principle for this was given in Gentzen [7]. The new feature of GL_{EF} is the interactions of thought sequents with different epistemic statuses. An epistemic structure E has the role to constrain the language, and F constrains interpersonal inferences.

A *proof* P of $\mathrm{B}_e[\Gamma \to \Theta]$ in GL_{EF} is a finite tree satisfying:

P1: a thought sequent admissible in E is associated with each node;

P2: the thought sequent associated with each leaf (endnode) is an instance of the axiom (initial sequent);

P3: adjoining nodes together with their associated thought sequents form an instance of one of the above inference rules;

P4: $\mathrm{B}_e[\Gamma \to \Theta]$ is associated with the root node;

P5: e' belongs to F for any thought sequent $\mathrm{B}_{e'}[\Delta \to \Lambda]$ in the proof P.

We say that $\mathrm{B}_e[\Gamma \to \Theta]$ is *provable* in GL_{EF}, denoted by $\vdash_{EF}\mathrm{B}_e[\Gamma \to \Theta]$, iff there is a proof P of $\mathrm{B}_e[\Gamma \to \Theta]$ in GL_{EF}. We write $\nvdash_{EF}\mathrm{B}_e[\Gamma \to \Theta]$ for the negation of $\vdash_{EF}\mathrm{B}_e[\Gamma \to \Theta]$.

When $E = F = \{\epsilon\}$, GL_{EF} is *classical propositional logic*, which is denoted by GL_0 with its provability relation $\vdash_0$.

When we do not restrict E and F, that is, $E = F = N^{<\omega}$, the resulting logic is denoted by GL, and its provability is denoted by $\vdash$. In this logic, even the outer $\mathrm{B}_e[\cdots]$ becomes unnecessary. Thus, it suffices to consider the provability of a sequent $\Gamma \to \Theta$. This system is called KD^3 in the literature. The Hilbert-style counterpart of KD^3 is defined from the classical logic by adding Axiom K: $\mathrm{B}_i(A \supset C) \supset (\mathrm{B}_i(A) \supset \mathrm{B}_i(C))$, and Axiom D: $\neg\mathrm{B}_i(\neg A \wedge A)$, and the Necessitation Rule: $\frac{A}{\mathrm{B}_i(A)}$.

In the above system, epistemic axioms such as Truthfulness: $\mathrm{B}_i(A) \supset A$, Positive Introspection: $\mathrm{B}_i(A) \supset \mathrm{B}_i\mathrm{B}_i(A)$, and Negative Introspection: $\neg\mathrm{B}_i(A) \supset \mathrm{B}_i(\neg\mathrm{B}_i(A))$, are not assumed. In our resolution of the puzzle, those axioms are unnecessary. Besides this point, we will show that the epistemic depths required for our resolution can be very shallow.

We describe the 3 child muddy children puzzle in GL_{EF} with some epistemic structures E and F. We can describe the reasonings by the 3 children together in one logic GL_{EF}. In this case, E and F include epistemic statuses of the form $(i_1, ..., i_m)$ with $i_1 = 1, 2, 3$, i.e, E and F include epistemic statuses in the mind of each child. In fact, we can separate the minds of the 3 children: We define, for $i = 1, 2, 3$,

$$\begin{aligned} E_i &= \{(i_1, ..., i_m) \in E : i_1 = i\} \cup \{\epsilon\} \text{ and} \\ F_i &= \{(i_1, ..., i_m) \in F : i_1 = i\} \cup \{\epsilon\}. \end{aligned} \tag{2.4}$$

Of course, E (and F) is the union of E_1, E_2, E_3 (F_1, F_2, F_3), which are each epistemic structures themselves. Let $\Gamma_1, \Gamma_2, \Gamma_3$ be sets of formulas for $i = 1, 2, 3$ so that each formula in Γ_i has the outermost $\mathrm{B}_i(\cdot)$, and let A_1, A_2, A_3 be formulas. It was shown in Kaneko-Suzuki [15] (Theorem 3.9) that:

$$\vdash_{EF}\Gamma_1, \Gamma_2, \Gamma_3 \to \wedge\{\mathrm{B}_i(A_i)\}_{i=1,2,3}$$

$$\text{if and only if}$$

$$\vdash_{E_iF_i}\Gamma_i \to \mathrm{B}_i(A_i) \text{ for } i = 1, 2, 3.$$

It means that to have a statement about child i's provability, it is enough to concentrate on $\mathrm{GL}_{E_iF_i}$. In the following analysis, we focus on $\mathrm{GL}_{E_iF_i}$.

3. Muddy Children Process and Resolution

In this section, we formalize the process of the muddy children puzzle and the notion of a resolution for it. Our interest is in analyzing the resolution and the necessity of various components.

First, we describe a child's beliefs on the general background. These beliefs describe the environment, and are constant for all stages of the process. We fix a child i and consider one order of the other two children as (j,k). We may use other variables s and s' for names including i. We take the following *candidates of basic beliefs* of child i, based on the order (j,k), though some other beliefs may also be available:

(**Announcement by teacher**) Child i has beliefs of the announcement at three epistemic levels:

(a1): $\mathrm{B}_i(m_i \vee m_j \vee m_k)$; $\mathrm{B}_{(i,j)}(m_i \vee m_j \vee m_k)$; $\mathrm{B}_{(i,j,k)}(m_i \vee m_j \vee m_k)$.

(**Observational ability**) Child i believes the following observational abilities:

(a2): $\mathrm{B}_i(\neg m_{s'} \supset \mathrm{B}_s(\neg m_{s'}))$ for $s \neq s'$; $\mathrm{B}_{(i,j)}(\neg m_{s'} \supset \mathrm{B}_s(\neg m_{s'}))$ for $s \neq s'$;

(**Action rule)** Child i has beliefs of the action rules at two epistemic levels:

(a3): $\mathrm{B}_i(\mathrm{B}_s(m_s) \supset r_s)$ for $s = i,j,k$; $\mathrm{B}_{(i,j)}(\mathrm{B}_s(m_s) \supset r_s)$ for $s = i,j,k$.[3]

The beliefs expressed in (a1) are about the announcement by the teacher that at least one face is muddy. Child i sees that the others hear the announcement, but we take a specific interpersonal order of these observations. The deepest belief about the announcement is $\mathrm{B}_{(i,j,k)}(m_i \vee m_j \vee m_k)$, which is expressed in the picture of Figure 3.1. The beliefs expressed in (a2) are about the depth of vision of children regarding the faces of others, in particular "if a child is not muddy, then the other children see that he is not muddy". The beliefs expressed in (a3) are about the action rule which states that "if a child believes he is muddy, then he raises his hand".

In fact, as mentioned already, the above list is incomplete and we have only taken some subset of all the basic beliefs that might be available of child i. Our purpose is to take a large enough set to obtain a resolution, and to find which of those beliefs are necessary for a resolution. We denote the set of beliefs listed above by Γ_i^0, and we will consider a subset of Γ_i^0 as possibly part of a resolution.

By the nature of the puzzle, some beliefs are coming as time proceeds. In particular, the responses to the teacher's question by the other children provide a source for new beliefs. This process may have the stages $t = 1, 2, ...$ Stage 1 is the start, and each child thinks about the teacher's question based on his own basic beliefs in the beginning of stage 1. The general description of a stage t is given as follows.

Stage t: The stage consists of:

(**Teacher's Question)** At the beginning of the stage, the teacher asks: "Do you know your face is muddy? If so, raise your hand."

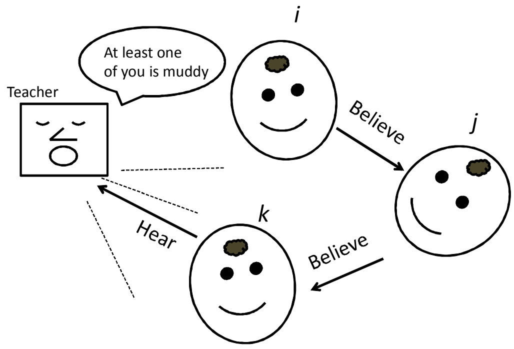
[3] The author thanks Tai-Wei Hu and a referee for suggesting this form of action rules connecting beliefs and actions.

(**Inferences**) Next, each child analyzes his beliefs in order to answer the teacher's question.

(**Actions)** After his analysis, each child chooses an action to raise or not raise his hand.

(**Observations**) Finally, at the end of the stage, each observes the actions of the others. These observations may be transformed into new beliefs for the next stage.

To discuss a resolution for child i, we fix the logic $\mathrm{GL}_{E_iF_i}$ for child i. In the following, we adopt the epistemic structures E_i and F_i in the form of (2.4). A resolution will consist of a finite sequence $\langle \Gamma_i^1, ..., \Gamma_i^\ell \rangle$ of beliefs of child i starting with his basic beliefs Γ_i^1 and finishing with his beliefs Γ_i^ℓ at stage ℓ where he proves his face is muddy. We take E_i big enough so that each sequent $\Gamma_i^t \rightarrow \mathrm{B}_i(m_i)$ is admissible in E_i for each $t = 1, ..., \ell$.



$\mathrm{B}_{(i,j,k)}(m_i \vee m_j \vee m_k)$

Figure 3.1: *i* believes *j* believes *k* hears the teacher's announcement

The process terminates after, and only after, a child raises his hand. In any intermediate stage, the only possible observations are the same, i.e., no child raises a hand. Consequently, only the observation of inaction needs to be considered for transitions between stages. We consider these observations to depth 2 for child i:

(Observations of Inactions of others)

(b): $\mathrm{B}_i(\neg r_s)$ for $s \neq i$; $\mathrm{B}_{(i,j)}(\neg r_s)$ for $s \neq j$.[4]

We denote the set of formulas in (b) by Δ_i. The transition of beliefs between stage $t-1$ and stage t is paramount to our analysis. At the end of stage $t-1$ where no child raises a hand, child i adds some belief A_i^{t-1} from Δ_i to Γ_i^{t-1} to obtain the new belief set $\Gamma_i^t = \Gamma_i^{t-1} \cup \{A_i^{t-1}\}$ for stage t. This process continues until someone raises his hand.

[4]We can allow these for all s, but the proofs of Theorems 3.2 and 3.3 become longer. The intent is that each uses his observations of others inactions, not his own actions to make inferences.

We say that a finite sequence $\langle \Gamma_i^1, ..., \Gamma_i^\ell \rangle$ is a *resolution* iff:

R1: $\Gamma_i^1 \subseteq \Gamma_i^0$;

R2: for $t > 1$, $\Gamma_i^t = \Gamma_i^{t-1} \cup \{A_i^{t-1}\}$ and $A_i^{t-1} \in \Delta_i$;

R3: $\delta^*(A_i^{t-1}) \supseteq \delta^*(A_i^t)$ for $2 \leq t < \ell$, where $\delta^*(A)$ is the smallest superset of $\delta(A)$ satisfying (2.1);

R4: for $t < \ell$, $\nvdash_{E_iF_i} \Gamma_i^t \rightarrow \mathrm{B}_i(m_i)$;

R5: $\vdash_{E_iF_i} \Gamma_i^\ell \rightarrow \mathrm{B}_i(m_i)$,

A resolution for child i describes his beliefs at each stage, and how these beliefs evolve over time. R1 restricts the initial beliefs Γ_i^1 to be a subset of the basic constant beliefs listed in Γ_i^0. R2 and R3 restrict the addition of new beliefs over time.

R2 has two parts: (i) some addition of new beliefs is required and (ii) it is a single belief at each stage. (i) adds some observations about inactions of other children, but without this, the situation for him is the same as the previous stage and he cannot make progress. (ii) and R3 are based on the basic principle that each child needs one stage to analyze a current observation of some inaction. This basic principle is applied even to the children in the mind of child i. Since these are subtle, we give more detailed comments on them after Theorem 3.1.

R4 implies that the process does not stop before a child finds his face is muddy. R5 states that child i finds his face is muddy at stage ℓ and raises his hand.

Now, we give a specific resolution with length $\ell = 3$. This theorem follows from Theorem 4.1, the proof of which is given in Section 4. Proofs of the other theorems of Section 3 will be given in Section 7.

Theorem 3.1 (Existence with Length $\ell = 3$): The sequence $\langle \Gamma_i^1, \Gamma_i^2, \Gamma_i^3 \rangle$ defined by the following (a), (b), and (c) is a resolution:

(a) (**Basic Beliefs**) Γ_i^1 is given as:

(a1) $\mathrm{B}_{(i,j,k)}(m_i \vee m_j \vee m_k)$,

(a2) $\mathrm{B}_i(\neg m_i \supset \mathrm{B}_j(\neg m_i))$, $\mathrm{B}_{(i,j)}(\neg m_i \supset \mathrm{B}_k(\neg m_i))$, $\mathrm{B}_{(i,j)}(\neg m_j \supset \mathrm{B}_k(\neg m_j))$,

(a3) $\mathrm{B}_i(\mathrm{B}_j(m_j) \supset r_j)$, $\mathrm{B}_{(i,j)}(\mathrm{B}_k(m_k) \supset r_k)$.

(b) **(Observations of inaction)** $A_i^1 = \mathrm{B}_{(i,j)}(\neg r_k)$ and $A_i^2 = \mathrm{B}_i(\neg r_j)$;

(c) **(Epistemic Structures)** $E_i = F_i = \{(i,j,k), (i,j), (i), \epsilon\}$.

The resolution given in Theorem 3.1 turns out to be the smallest, which will be shown in Theorems 3.2 and 3.3. Here, we look at this resolution in more detail.

First, this resolution gives a set of basic beliefs Γ_i^1. The only belief about the announcement in Γ_i^1 is $\mathrm{B}_{(i,j,k)}(m_i \vee m_j \vee m_k)$. The other beliefs included in Γ_i^1 are about observations of faces and the action rules. These are only a subset of the full set of constant basic beliefs Γ_i^0.

A natural question is why the belief $\mathrm{B}_i(m_j \wedge m_k)$ (child i sees the faces of j and k) is not in the list of basic beliefs in Theorem 3.1. Indeed, it would be reasonable to include this belief as part of the basic beliefs. These types of beliefs are, in fact, partially included in (a2), e.g., the first belief $\mathrm{B}_i(\neg m_i \supset \mathrm{B}_j(\neg m_i))$ comes from the fact that each child believes that each of the three sees the faces of the others. For the resolution, however, we need only what (a2) includes and the other beliefs like $\mathrm{B}_i(m_j \wedge m_k)$ can be excluded.

In (b) of Theorem 3.1, the new beliefs of child i coming after the inactions of others at stages 1 and 2 are listed. In the transition from stage 1 to 2, child i adds the belief $\mathrm{B}_{(i,j)}(\neg r_k)$ to obtain Γ_i^2. This is used in the analysis part of stage 2 by child i to infer that $\mathrm{B}_{(i,j)}(m_i \vee m_j)$. Since i cannot yet reach the conclusion $\mathrm{B}_i(m_i)$, however, he does not raise his hand in stage 2. At the end of stage 2, he observes that no other child raises his hand either. Next, in the transition from stage 2 to stage 3, child i adds the belief $\mathrm{B}_i(\neg r_j)$ to obtain Γ_i^3. In the analysis part of stage 3, he uses $\mathrm{B}_i(\neg r_j)$, together with his previous inference $\mathrm{B}_{(i,j)}(m_i \vee m_j)$ from stage 2, to conclude that $\mathrm{B}_i(m_i)$. Thus, he raises his hand in stage 3.

Now we return to the comments on R2(ii) and R3. We ordered the children i, j, k so that i considers j directly, and k indirectly through j's mind. R3 requires child i to first think about j's observation of k's inaction in one stage, and then for child i to look at j's response to it in a later stage. The order should not be mixed, which is R3, and they should not be taken at the same stage, which is R2(ii).

A counter-example violating R2(ii) happens when child i take the two new beliefs $\mathrm{B}_{(i,j)}(\neg r_k)$ and $\mathrm{B}_i(\neg r_j)$ at stage 2. Then, he can prove his face is muddy already in stage 2. However, the addition of $\mathrm{B}_i(\neg r_j)$ does not give the time for j to respond to his observation of k's inaction. This violates the basic principle that child i needs to wait for j's response to k's inaction before using $\mathrm{B}_i(\neg r_j)$.

A counter-example violating R3 happens when i mixes the order and takes $\mathrm{B}_i(\neg r_j)$ first in stage 2, followed by $\mathrm{B}_{(i,j)}(\neg r_k)$ in stage 3. Then, in stage 2, he cannot progress at all. If, however, he uses both beliefs in stage 3, then he can conclude his face is muddy. But this also violates the basic principle that the sudden use of $\mathrm{B}_i(\neg r_j)$ after $\mathrm{B}_{(i,j)}(\neg r_k)$ does not give the time for j to respond to his observation of k's inaction.

In order to make the above basic principle more explicit, we may need to distinguish the beliefs adopted at different times. This may enable us to have more meaningful restrictions on the order of adoption of new beliefs and reasoning. In the present paper, however, we have embodied this principle in R2(ii) and R3.

We may have other resolutions. However, the next theorem states that there is no shorter one.

Theorem 3.2 (No Resolution for $\ell < 3$) If $\langle \Gamma_i^1, ..., \Gamma_i^\ell \rangle$ is a resolution, then $\ell \geq 3$.

By Theorems 3.1 and 3.2, the shortest resolution has $\ell = 3$ and Theorem 3.1 gives an example. We may wonder if the beliefs or epistemic structures in Theorem 3.1 can be reduced, i.e., if there is a resolution with a smaller belief set, or shallower epistemic structures. The next theorem shows that the resolution given by Theorem 3.1 is the smallest.

Theorem 3.3 (Minimum Resolution): If $\langle \Gamma_i^1, \Gamma_i^2, \Gamma_i^3 \rangle$ is a resolution, then:
(a) Γ_i^1 contains the beliefs in (a1), (a2), (a3) of Theorem 3.1;
(b) A_i^1, A_i^2 are given by (b) of Theorem 3.1;
(c) F_i includes the epistemic structures of Theorem 3.1, i.e., $F_i \supseteq \{(i, j, k), (i, j), (i), \epsilon\}$.

Part (a) states that each resolution of length $\ell = 3$ includes all the basic beliefs given in Theorem 3.1(a), though it may have some more. Part (b) states that each resolution has the added observations listed in the order given in Theorem 3.1(b). Part (c) states that the epistemic structures given in Theorem 3.1(c) are the shallowest for a resolution.

4. Extension to n children

We show here how the result of Theorem 3.1 can be extended to n children. As might be expected, the length of the resolution is n, and the interpersonal beliefs are up to depth n.

Consider a number $n \geq 1$ of muddy children. We denote $\{1, ..., n\}$ by N, and use the notation $e_v = (1, ..., v)$ for $v \in N$. We will focus on the resolution for child 1 taking the fixed order $(1, 2, ..., n)$ of the children. Now, the language of GL_{EF} is constructed in the same way as in Section 2 from the primitive propositional symbols m_v and r_v for all $v \in N$.

The candidates of basic beliefs $\Gamma_1^0(n)$ are described by:
(Announcement): $\mathrm{B}_{e_v}(m_1 \vee \cdots \vee m_n)$ for all $v \leq n$;
(Observational Abilities): $\{\mathrm{B}_{e_v}(\neg m_{s'} \supset \mathrm{B}_s(\neg m_{s'})) : v < n \text{ and } s \neq s'\}$;
(Action Rules): $\{\mathrm{B}_{e_v}(\mathrm{B}_s(m_s) \supset r_s) : v < n \text{ and } s \in N\}$.
These generalize the sets of candidates of basic beliefs Γ_1^0 for child 1 of the 3 child case given in Section 3 with the order of children $(i, j, k) = (1, 2, 3)$. We define the set of candidate new beliefs possibly obtained at future stages as:

$$\Delta_1(n) = \{\mathrm{B}_{e_v}(\neg r_s) : v < n \text{ and } s \neq v\} \quad (4.1)$$

A resolution for child 1 in the n child case is defined in the same way as in Section 3 replacing Γ_1^0 by $\Gamma_1^0(n)$ and Δ_1 by $\Delta_1(n)$.

The specific resolution $\langle \Gamma_1^1(n), ..., \Gamma_1^n(n) \rangle$ for the n child case corresponding to the one given in Theorem 3.1 is given as follows: $\Gamma_1^1(n)$ consists of the following formulas:
(a1^n) $\mathrm{B}_{e_n}(m_1 \vee \cdots \vee m_n)$;
(a2^n) $\mathrm{B}_{e_v}(\neg m_s \supset \mathrm{B}_{v+1}(\neg m_s)), v < n$ and $s \leq v$;
(a3^n) $\mathrm{B}_{e_v}(\mathrm{B}_{v+1}(m_{v+1}) \supset r_{v+1}), v < n$.

Let $A_1^k(n) = \mathrm{B}_{e_{n-k}}(\neg r_{n+1-k})$ for $k < n$. For all $t \geq 2$, we define:

(b^n) $\Gamma_1^t(n) = \Gamma_1^1(n) \cup \{A_1^1(n), ..., A_1^{t-1}(n)\}$.
(c^n) $E_1(n) = F_1(n)$ is the smallest epistemic structure containing $(1, ..., n)$.

We have the following theorem.

Theorem 4.1 The sequence $\langle \Gamma_1^1(n), ..., \Gamma_1^n(n) \rangle$ defined by (a1^n), (a2^n), (a3^n), (b^n), (c^n) is a resolution.

It can be verified by inspection that the target sequence $\langle \Gamma_1^1(n), ..., \Gamma_n^n(n) \rangle$ satisfies R1, R2, R3. To prove Theorem 4.1, it remains to show R4 and R5. We first prove R5, which is Lemma 4.2. Then, we prove R4.

For R5, we must prove:

$$\vdash_{E_1(n)F_1(n)} \Gamma_1^n(n) \longrightarrow \mathrm{B}_1(m_1) \tag{4.2}$$

Our proof is by induction over the number of children $n \geq 1$. The definition of the final stage beliefs $\Gamma_1^n(n)$ is given without reference to any other case. However, by observation of the beliefs given in (a1^n), (a2^n), (a3^n), and (b^n) we find a connection between $\Gamma_1^n(n)$ and $\Gamma_1^{n+1}(n+1)$ in that:

$$\Gamma_1^n(n) \subseteq \Gamma_1^{n+1}(n+1) \cup \mathrm{B}_{e_n}(m_1 \vee \cdots \vee m_n) \tag{4.3}$$

The connection described in (4.3) will be used in our inductive step. The main argument in the inductive step is to reduce the $n+1$ child case to the n child case after the first observation of inactions by the others. For this, we will use the following lemma.

Lemma 4.1: Let $n \geq 1$. Then $\vdash_{E_1(n+1)F_1(n+1)} \Gamma_1^2(n+1) \longrightarrow \mathrm{B}_{e_n}(m_1 \vee \cdots \vee m_n)$.

We will give the proof of Lemma 4.1 after stating and proving the next lemma which proves R5.

Lemma 4.2: Let $n \geq 1$. Then (4.2) holds, i.e., $\vdash_{E_1(n)F_1(n)} \Gamma_1^n(n) \longrightarrow \mathrm{B}_1(m_1)$.

Proof. As mentioned above, we prove the theorem by induction over the number of children n. The base case is $n = 1$. Since $\mathrm{B}_1(m_1) \in \Gamma_1^1(1)$, (4.2) holds trivially for $n = 1$. The inductive hypothesis is that $\vdash_{E_1(k)F_1(k)} \Gamma_1^k(k) \longrightarrow \mathrm{B}_1(m_1)$ holds for some $k \geq 1$.

By the inductive hypothesis and $E_1(k) = F_1(k) \subseteq F_1(k+1) = E_1(k+1)$, we have:

$$\vdash_{E_1(k+1)F_1(k+1)} \Gamma_1^k(k) \longrightarrow \mathrm{B}_1(m_1). \tag{4.4}$$

Thinning the antecedent in (4.4) and using (4.3) we obtain:

$$\vdash_{E_1(k+1)F_1(k+1)} \Gamma_1^{k+1}(k+1), \mathrm{B}_{e_k}(m_1 \vee \cdots \vee m_k) \longrightarrow \mathrm{B}_1(m_1). \tag{4.5}$$

From Lemma 4.1, we have the sequent:

$$\vdash_{E_1(k+1)F_1(k+1)} \Gamma_1^2(k+1) \longrightarrow \mathrm{B}_{e_k}(m_1 \vee \cdots \vee m_k). \tag{4.6}$$

Since $\Gamma_1^2(k+1) \subseteq \Gamma_1^{k+1}(k+1)$, thinning the antecedent of (4.6) gives us:

$$\vdash_{E_1(k+1)F_1(k+1)} \Gamma_1^{k+1}(k+1) \longrightarrow \mathrm{B}_{e_k}(m_1 \vee \cdots \vee m_k). \tag{4.7}$$

By cut applied to (4.5) and (4.7) we obtain the sequent:

$$\vdash_{E_1(k+1)F_1(k+1)} \Gamma_1^{k+1}(k+1) \longrightarrow \mathrm{B}_1(m_1). \tag{4.8}$$

Since k was chosen arbitrarily, (4.2) follows by the principle of induction. □

Proof of Lemma 4.1. The following sequents are provable in $\mathrm{GL}_{E_1(n+1)F_1(n+1)}$:
(a) $\vdash_{E_1(n+1)F_1(n+1)} \mathrm{B}_{e_{n+1}} [(m_1 \vee \cdots \vee m_{n+1}), \{\neg m_s\}_{s \neq n+1} \ \rightarrow \ m_{n+1}]$;
(b) $\vdash_{E_1(n+1)F_1(n+1)} \mathrm{B}_{e_n} [\ \ \rightarrow \ (m_1 \vee \cdots \vee m_n), \neg m_{s'}]$ for $s' \neq n+1$;
(c) $\vdash_{E_1(n+1)F_1(n+1)} \mathrm{B}_{e_n} [\mathrm{B}_{n+1}(m_{n+1}) \supset r_{n+1},\ \neg r_{n+1} \rightarrow \neg \mathrm{B}_{n+1}(m_{n+1})]$.
(d) $\vdash_{E_1(n+1)F_1(n+1)} \mathrm{B}_{e_n} [\mathrm{B}_{n+1}(m_1 \vee \cdots \vee m_{n+1}), \{\neg m_s \supset \mathrm{B}_{n+1}(\neg m_s)\}_{s \neq n+1}, \neg \mathrm{B}_{n+1}(m_{n+1}) \rightarrow (m_1 \vee \cdots \vee m_n)]$;
Statements (a), (b), and (c) are straightforward. Applying ($\mathrm{B}_{n+1} \rightarrow \mathrm{B}_{n+1}$) to (a), and then ($\neg \rightarrow$) to the resulting sequent, we obtain:

$$\vdash_{E_1(n+1)F_1(n+1)} \mathrm{B}_{e_n} [\mathrm{B}_{n+1}(m_1 \vee \cdots \vee m_{n+1}),\ \{\mathrm{B}_{n+1}(\neg m_s)\}_{s \neq n+1},\ \neg \mathrm{B}_{n+1}(m_{n+1}) \rightarrow \ \ ] \tag{4.9}$$

Applying ($\supset\rightarrow$) to (4.9) and (b) for the cases $s' = 1, ..., n$ in succession, we obtain (d). Applying cut to (c) and (d), we obtain:

$$\vdash_{E_1(n+1)F_1(n+1)} \mathrm{B}_{e_n} [\mathrm{B}_{n+1}(m_1 \vee \cdots \vee m_{n+1}), \{\neg m_s \supset \mathrm{B}_{n+1}(\neg m_s)\}_{s \neq n+1}, \mathrm{B}_{n+1}(m_{n+1}) \supset r_{n+1},\ \neg r_{n+1}, \rightarrow \ (m_1 \vee \cdots \vee m_n)]. \tag{4.10}$$

Applying ($\mathrm{B}_i \rightarrow \mathrm{B}_i$) to this sequent, for the cases $i = n$,...,1 in succession we obtain:

$$\vdash_{E_1(n+1)F_1(n+1)} [\mathrm{B}_{e_{n+1}}(m_1 \vee \cdots \vee m_{n+1}), \{\mathrm{B}_{e_n}(\neg m_s \supset \mathrm{B}_{n+1}(\neg m_s)\}_{s \neq n+1}, \mathrm{B}_{e_n}(\mathrm{B}_{n+1}(m_{n+1}) \supset r_{n+1}), \mathrm{B}_{e_n}(\neg r_{n+1}) \rightarrow \ \mathrm{B}_{e_n}(m_1 \vee \cdots \vee m_n)]. \tag{4.11}$$

Since all the beliefs in the antecedent of (4.11) are contained in $\Gamma_1^2(n+1)$, the sequence $\vdash_{E_1(n+1)F_1(n+1)} \Gamma_1^2(n+1) \longrightarrow \mathrm{B}_{e_n}(m_1 \vee \cdots \vee m_n)$ is obtained from (4.11) by thinning. $\square$

It remains only to show R4.

Proof of R4: Let $n \geq 1$. We need to show that $\nvdash_{E_1(n)F_1(n)} \Gamma_i^t(n) \rightarrow \mathrm{B}_i(m_i)$ for $t < n$. If $n = 1$, there is no $t < n$, so we presume in what follows that $n \geq 2$. Since $\Gamma_1^{t-1}(n) \subseteq \Gamma_1^t(n)$ for $t = 1, ..., n-1$, it suffices to show $\nvdash_{E_1(n)F_1(n)} \Gamma_1^{n-1}(n) \rightarrow \mathrm{B}_1(m_i)$. By Theorem 6.1, if we can show $\nvdash_0 \varepsilon_0\Gamma_1^{n-1}(n) \rightarrow \varepsilon_0\mathrm{B}_1(m_1)$, then we will have $\nvdash_{E_1(n)F_1(n)} \Gamma_1^{n-1}(n) \rightarrow \mathrm{B}_1(m_1)$, *a fortiori*, $\nvdash_{E_1(n)F_1(n)} \Gamma_1^t(n) \rightarrow \mathrm{B}_1(m_1)$ for $t < n$. Observe that $\varepsilon_0\Gamma_1^{n-1}(n)$ is a subset of:

$$\{(m_1 \vee \cdots \vee m_n)\} \cup \{(\neg m_s \supset \neg m_s), (m_s \supset r_s)\}_{s \in N} \cup \{\neg r_s\}_{s=3,\ldots,n}. \tag{4.12}$$

From this assumption set, we can obtain $m_1 \vee m_2$, but we cannot get to m_1. Hence, we conclude that$\nvdash_{E_1(n)F_1(n)} \Gamma_i^t(n) \rightarrow \mathrm{B}_i(m_i)$ for $t < n$. $\square$

The extensions of Theorem 3.2 and Theorem 3.3 to the n child case should also be possible, but the proofs require more complicated arguments than the ones used in the proofs given in Section 7.

5. Conclusions

In Section 1, we addressed the issue that the logical inferences are all implicit and informal in the information partition approach, and are still indirect in the semantical approach. In our approach, all logical inferences are explicitly described as classical inferences within some epistemic depth, though we sometimes move from a deeper to a shallower depth with the epistemic inference rule. The only elements that are additional to such classical inferences are the new beliefs about the observations of inactions of others.

In our approach, epistemic axioms such as Truthfulness, Positive Introspection, and Negative Introspection are not used at all. The semantic approach of van Ditmarsch, *et al.* [18] used models based on those additional axioms. Gerbrandy-Groeneveld [8] did not use those axioms, but they gave additional axioms to describe updating and assumed common knowledge on various components of their model.

Although we used some semantical methods in our analysis, the description of the puzzle itself is syntactical and proof theoretical. This enables us to avoid implicit assumptions sneaking into the system. Our analysis revealed what components are necessary for the resolution of the puzzle.

Now we enter a more detailed discussion about the results. Our Theorem 3.1 gives precise beliefs for a resolution involving epistemic depths only up to 3. In the resolution for child i, in stage 2, child j in the mind of child i uses $\neg r_k$ to infer that either j or i is muddy. Next, in stage 3, child i infers, from $\neg r_j$, that his own face is muddy. Here we see the separation of epistemic inferences at work.

As we also mentioned in the introduction, our aim was not just to give some precise beliefs and structure to the resolution, but also to analyze it in terms of the necessary components. The two other theorems of Section 3 give results about this. Theorem 3.2 states that the minimal length of a resolution is 3. Theorem 3.3 gives minimality in terms of the set of basic beliefs, the beliefs about inactions of the others, and the epistemic structure. It states that reducing any of these will prevent a resolution. This analysis has been successful in showing the precise components needed for a resolution to the muddy children puzzle for the 3 child case.

In Section 4 we showed how the analysis can be extended to the $n$child case and we gave an n child resolution in Theorem 4.1.

6. Two Meta-Theorems

In order to prove the theorems of the Section 3, will make use of two meta-theorems for logic GL_{EF}. These theorems are not original and are found in various places in the literature which we mention after each theorem.

The first meta-theorem is called the belief eraser theorem. Let Γ be a set of formulas. Define ε_0 to be the operator that removes all occurrences of B_1, B_2, B_3. For example if $\Gamma = \{\mathrm{B}_1(m_1) \supset m_1,\ \mathrm{B}_2\mathrm{B}_1(m_1 \vee m_2)\}$, then $\varepsilon_0\Gamma = \{m_1 \supset m_1, m_1 \vee m_2\}$. The belief eraser theorem states that the unprovability of some sequent $\varepsilon_0\Gamma \to \varepsilon_0\Theta$ in classical logic implies the unprovability of the thought sequent $\mathrm{B}_e[\Gamma \to \Theta]$ in the logic GL_{EF}.

Theorem 6.1 (**Belief Eraser**) Let Γ and Θ be finite subsets of $\mathcal{P}$ and let E and F be epistemic structures. If $\nvdash_0 \varepsilon_0\Gamma \to \varepsilon_0\Theta$, then $\nvdash_{EF}\mathrm{B}_e[\Gamma \to \Theta]$.

This theorem was given in Kaneko-Nagashima [11] for a different logical system. Nevertheless, it holds for the current system used in this paper and the proof follows the proof given there. Theorem 6.1 will be used to show some unprovability results in Theorems 3.1 and 3.2.

Our second meta-theorem is soundness and completeness of the epistemic logic GL. Although we use the epistemic logic of shallow depths GL_{EF}, it is enough for our aims to refer to the semantics[5] for GL.

We follow the presentation of soundness and completeness given in Kaneko [10]. We say that $\mathcal{K} = (W, R_1, R_2, R_3)$ is a *Kripke Frame* iff W is a non-empty set, and R_i is a binary relation on W for each $i = 1, 2, 3$. An *assignment* σ in a Kripke Frame $\mathcal{K}$ is a function from $\{m_1, m_2, m_3\}$ to $\{\bot, \top\}$. We call a pair $(\mathcal{K}, \sigma)$ a *Kripke model*. We

[5] The semantics for GL_{EF} is developed Kaneko-Suzuki [13].

define the *valuation relation* $(\mathcal{K}, \sigma, w) \models$ and its negation $(\mathcal{K}, \sigma, w) \not\models$ for each $w \in W$ by induction on the length of a formula:

K0: for each $p \in \{m_1, m_2, m_3\}$, $(\mathcal{K}, \sigma, w) \models p$ iff $\sigma(w, p) = \top$;

K1: $(\mathcal{K}, \sigma, w) \models \neg A$ iff $(\mathcal{K}, \sigma, w) \not\models A$;

K2: $(\mathcal{K}, \sigma, w) \models A \supset B$ iff $(\mathcal{K}, \sigma, w) \models \neg A$ or $(\mathcal{K}, \sigma, w) \models B$;

K3: $(\mathcal{K}, \sigma, w) \models \wedge\Phi$ iff $(\mathcal{K}, \sigma, w) \models A$ for all A in Φ;

K4: $(\mathcal{K}, \sigma, w) \models \vee\Phi$ iff $(\mathcal{K}, \sigma, w) \models A$ for some A in Φ;

K5: $(\mathcal{K}, \sigma, w) \models \mathrm{B}_i(A)$ iff $(\mathcal{K}, \sigma, u) \models A$ for all u with $(w, u) \in R_i$.

We say that a Kripke frame $\mathcal{K} = (W, R_1, R_2, R_3)$ is *serial* iff for any $i = 1, 2, 3$ and any $w \in W$, there is some $u \in W$ such that $(w, u) \in R_i$. Seriality corresponds to the logic we are using. Recall that $\vdash \Gamma \to A$ means that $\Gamma \to A$ is provable in the logic GL. We have the following:

Theorem 6.2 (Soundness and Completeness) Let Γ be a finite nonempty subset of $\mathcal{P}$ and let A be a formula in $\mathcal{P}$. Then, $\vdash \Gamma \to A$ if and only if for all serial Kripke models $(\mathcal{K}, \sigma)$, $(\mathcal{K}, \sigma, w) \models \wedge\Gamma \supset A$.

Soundness and completeness theorems for various modal logics are given in Hughes-Cresswell [9]. Though they used Hilbert-style systems, their results can be translated to Gentzen-style systems including GL using the translation given in Kaneko-Nagashima [12], which proves Theorem 6.2.

We will use Theorem 6.2 to show unprovabilities of some statements in parts (b) and (c) of Theorem 3.3. It is used in the following way. Notice that it is the negative form of this theorem that $\nvdash \Gamma \to A$ if and only if there is a Kripke model $(\mathcal{K}, \sigma)$ and a $w \in W$ such that $(\mathcal{K}, \sigma, w) \not\models \wedge\Gamma \supset A$. Since E and F are restrictions on the epistemic logic GL, $\vdash_{EF} \Gamma \to A$ implies $\vdash \Gamma \to A$, equivalently, $\nvdash \Gamma \to A$ implies $\nvdash_{EF} \Gamma \to A$. Therefore, if we find a model $(\mathcal{K}, \sigma)$ and a $w \in W$ with $(\mathcal{K}, \sigma, w) \not\models \wedge\Gamma \supset A$, then by the negative form of the above theorem, $\nvdash \Gamma \to A$, which implies $\nvdash_{EF} \Gamma \to A$.

7. Proofs of the theorems

We now are in a position to prove the remaining two theorems of Section 3 about the 3 child puzzle. In Subsection 7.1 we prove Theorem 3.2. In Subsection 7.2 we prove Theorem 3.3.

7.1. Proof of Theorem 3.2

Suppose that $\langle \Gamma_i^1, ..., \Gamma_i^\ell \rangle$ satisfies R1 - R4 with $\ell < 3$. We will show that $\langle \Gamma_i^1, ..., \Gamma_i^\ell \rangle$ violates R5, i.e., it is not a resolution. Since $\ell < 3$, either $\ell = 1$ or $\ell = 2$.

Suppose $\ell = 1$. Since $\langle \Gamma_i^1 \rangle$ satisfies R1, $\Gamma_i^1 \subseteq \Gamma_i^0$. Hence, to prove R5 is violated, it suffices to show $\nvdash_{E_iF_i} \Gamma_i^0 \to \mathrm{B}_i(m_i)$. By Theorem 6.1, if we can show $\nvdash_0 \varepsilon_0\Gamma_i^0 \to \varepsilon_0\mathrm{B}_i(m_i)$, then $\nvdash_{E_iF_i} \Gamma_i^0 \to \mathrm{B}_i(m_i)$. Observe that $\varepsilon_0\Gamma_i^0 \to \varepsilon_0\mathrm{B}_i(m_i)$ is:

$$m_i \vee m_j \vee m_k,\ \{\neg m_s \supset \neg m_s,\ m_s \supset r_s\}_{s=i,j,k} \to m_i \tag{7.1}$$

Here we cannot eliminate m_k or m_j from $m_i \vee m_j \vee m_k$ using classical logic. Hence, $\nvdash_0 \varepsilon_0\Gamma_i^0 \to \varepsilon_0\mathrm{B}_i(m_i)$.

Suppose $\ell = 2$. Then, either $A_i^1 = \mathrm{B}_i(\neg r_s)$ for $s \neq i$, or $A_i^1 = \mathrm{B}_{(i,j)}(\neg r_s))$ for $s \neq j$. We treat the case of $A_i^1 = \mathrm{B}_i(\neg r_s)$ for $s \neq i$. The other case can be treated in a similar manner.

Let $\Gamma_i^2 = \Gamma_i^1 \cup \{\mathrm{B}_i(\neg r_s)\}$ for $s \neq i$. By R1 and Theorem 6.1, it suffices to show $\nvdash_0 \varepsilon_0\Gamma_i^0 \cup \{\varepsilon_0\mathrm{B}_i(\neg r_s)\} \to \varepsilon_0\mathrm{B}_i(m_i)$. Observe that $\varepsilon_0\Gamma_i^0 \cup \{\varepsilon_0\mathrm{B}_i(\neg r_s))\} \to \varepsilon_0\mathrm{B}_i(m_i)$ is:

$$m_i \vee m_j \vee m_k, \neg r_s,\ \{\neg m_{s'} \supset \neg m_{s'}, m_{s'} \supset r_{s'}\}_{s'=i,j,k} \to m_i. \tag{7.2}$$

Since $s \neq i$, there are two cases. The case of $s = k$ is similar to (7.1) with some additional part in the antecedent that does not help us to get to m_i. The remaining case is $s = j$, which differs from the case of $s = k$ only in the antecedent, where $\neg r_k$ is replaced by $\neg r_j$ in (7.2). Though $\neg r_j$, together with $m_j \supset r_j$ eliminates m_j from $m_i \vee m_j \vee m_k$, again we cannot get to m_i in classical logic. Hence, in each case, we conclude that $\nvdash_0 \varepsilon_0\Gamma_i^0 \cup \{\varepsilon_0\mathrm{B}_i(\neg r_s)\} \to \varepsilon_0\mathrm{B}_i(m_i)$. □

7.2. Proof of Theorem 3.3

This theorem consists of three statements (a), (b), and (c). Since the proof of (a) is long, and it does not need (b), we start with the proof of (b). Next, we prove (a), and finally we prove (c) ((c) needs (a) and (b)). Throughout the proof, we assume that $\langle \Gamma_i^1, \Gamma_i^2, \Gamma_i^3 \rangle$ satisfies R1, R2, and R3. We prove the contrapositive of each assertion, e.g., if (b) is violated, then $\langle \Gamma_i^1, \Gamma_i^2, \Gamma_i^3 \rangle$ is not a resolution, i.e., it violates R4 or R5. The proofs of (a) and (c) are done likewise.

Proof of (b): Suppose that (b) is violated, i.e., $A_i^1 \neq \mathrm{B}_{(i,j)}(\neg r_k))$ or $A_i^2 \neq \mathrm{B}_i(\neg r_j))$. By R2 and R3, $A_i^1, A_i^2 \in \Delta_i$ and $\delta^*(A_i^1) \supseteq \delta^*(A_i^2)$.

Suppose $A_i^1 = A_i^2$. Then $\Gamma_i^2 = \Gamma_i^3$ by R2. If R5 holds, i.e., $\vdash_{E_iF_i} \Gamma_i^3 \to \mathrm{B}_i(m_i)$, then $\vdash_{E_iF_i} \Gamma_i^2 \to \mathrm{B}_i(m_i)$, a violation of R4.

Let $A_i^1 \neq A_i^2$. Then by R3, $A_i^1 = \mathrm{B}_{(i,j)}(\neg r_s))$ for some $s \neq j$ and $A_i^2 = \mathrm{B}_i(\neg r_{s'})$ for some $s' \neq i$. By assumption that (b) is violated, $A_i^1 \neq \mathrm{B}_{(i,j)}(\neg r_k)$ or $A_i^2 \neq \mathrm{B}_i(\neg r_j))$. This leaves 3 cases:

(i) $A_i^1 = \mathrm{B}_{(i,j)}(\neg r_i)$ and $A_i^2 = \mathrm{B}_i(\neg r_j)$;

(ii) $A_i^1 = \mathrm{B}_{(i,j)}(\neg r_i)$ and $A_i^2 = \mathrm{B}_i(\neg r_k)$;

(iii) $A_i^1 = \mathrm{B}_{(i,j)}(\neg r_k)$ and $A_i^2 = \mathrm{B}_i(\neg r_k)$;

We treat case (i) and show that R5 is violated. The other two cases can be eliminated in a similar manner.

By R1 and Theorem 6.1, to show R5 is violated, it suffices to show $\nvdash_0 \ \varepsilon_0\Gamma_i^0 \cup \{\varepsilon_0\mathrm{B}_{(i,j)}(\neg r_i)\} \cup \{\varepsilon_0\mathrm{B}_i(\neg r_j)\} \to \varepsilon_0\mathrm{B}_i(m_i)$. Observe that $\varepsilon_0\Gamma_i^0 \cup \{\varepsilon_0\mathrm{B}_{(i,j)}(\neg r_i)\} \cup \{\varepsilon_0\mathrm{B}_i(\neg r_j)\} \to \varepsilon_0\mathrm{B}_i(m_i)$ is:

$$m_i \vee m_j \vee m_k, \neg r_i, \neg r_j, \{\neg m_s \supset \neg m_s, m_s \supset r_s\}_{s=i,j,k} \to m_i. \tag{7.3}$$

This sequent differs from (7.2) with $s = i$ only in that the antecedent contains both $\neg r_i$ and $\neg r_j$. However, this does not allow us to get to m_i in classical logic. Hence, we conclude $\nvdash_0 \ \varepsilon_0\Gamma_i^0 \cup \{\varepsilon_0\mathrm{B}_{(i,j)}(\neg r_i))\} \cup \{\varepsilon_0\mathrm{B}_i(\neg r_j))\} \to \varepsilon_0\mathrm{B}_i(m_i)$. □

Proof of (a): Suppose that $\langle \Gamma_i^1, \Gamma_i^2, \Gamma_i^3 \rangle$ satisfying (b) violates (a). It follows that one of the following holds:

(I) $\mathrm{B}_{(i,j,k)}(m_i \vee m_j \vee m_k) \notin \Gamma_i^1$;

(II) $\mathrm{B}_{(i,j)}(\neg m_j \supset \mathrm{B}_k(\neg m_j)) \notin \Gamma_i^1$;

(III) $\mathrm{B}_{(i,j)}(\neg m_i \supset \mathrm{B}_k(\neg m_i)) \notin \Gamma_i^1$;

(IV) $\mathrm{B}_i(\neg m_i \supset \mathrm{B}_k(\neg m_i)) \notin \Gamma_i^1$;

(V) $\mathrm{B}_{(i,j)}(\mathrm{B}_k(m_k) \supset r_k) \notin \Gamma_i^1$;

(VI) $\mathrm{B}_i(\mathrm{B}_j(m_j) \supset r_j) \notin \Gamma_i^1$.

By R1 and (b), we have $\Gamma_i^1 \subseteq \Gamma_i^0$, $A_i^1 = \mathrm{B}_{(i,j)}(\neg r_k)$, and $A_i^2 = \mathrm{B}_i(\neg r_j)$.

Suppose, e.g., that (I) holds. Consider

$$\tilde{\Gamma}_i^3(\mathrm{I}) = \Gamma_i^0 \cup \{\mathrm{B}_{(i,j)}(\neg r_k), \mathrm{B}_i(\neg r_j)\} - \{\mathrm{B}_{(i,j,k)}(m_i \vee m_j \vee m_k)\}. \tag{7.4}$$

We shall prove $\nvdash_{E_iF_i} \tilde{\Gamma}_i^3(\mathrm{I}) \to \mathrm{B}_i(m_i)$, which implies $\nvdash_{E_iF_i} \Gamma_i^3 \to \mathrm{B}_i(m_i)$ since $\Gamma_i^3 \subseteq \tilde{\Gamma}_i^3(\mathrm{I})$. For this unprovability, we shall construct a serial Kripke model $(\mathcal{K}, \sigma)$ and a world $w \in W$ such that $(\mathcal{K}, \sigma, w) \nvDash \wedge\tilde{\Gamma}_i^3(\mathrm{I}) \supset \mathrm{B}_i(m_i)$. By Theorem 6.2 and the remark after it, we have $\nvdash \tilde{\Gamma}_i^3(\mathrm{I}) \to \mathrm{B}_i(m_i)$, and thus, $\nvdash_{E_iF_i} \tilde{\Gamma}_i^3(\mathrm{I}) \to \mathrm{B}_i(m_i)$, *a fortiori*, $\nvdash_{E_iF_i} \Gamma_i^3 \to \mathrm{B}_i(m_i)$, a violation of R5.

We construct a serial Kripke model for each case of (I) - (VI) that has the properties mentioned in the previous paragraph. Observe that the set $\Gamma_i^0 \cup \{\mathrm{B}_{(i,j)}(\neg r_k), \mathrm{B}_i(\neg r_j)\}$ consists of the following 23 formulas which we call *the list*:

1. $\mathrm{B}_i(m_i \vee m_j \vee m_k)$; 2. $\mathrm{B}_{(i,j)}(m_i \vee m_j \vee m_k)$; 3. $\mathrm{B}_{(i,j,k)}(m_i \vee m_j \vee m_k)$;
4. $\mathrm{B}_i(\neg m_i \supset \mathrm{B}_j(\neg m_i))$; 5. $\mathrm{B}_i(\neg m_i \supset \mathrm{B}_k(\neg m_i))$; 6. $\mathrm{B}_{(i,j)}(\neg m_i \supset \mathrm{B}_j(\neg m_i))$;
7. $\mathrm{B}_{(i,j)}(\neg m_i \supset \mathrm{B}_k(\neg m_i))$; 8. $\mathrm{B}_{(i,j)}(\neg m_j \supset \mathrm{B}_i(\neg m_j))$; 9. $\mathrm{B}_{(i,j)}(\neg m_j \supset \mathrm{B}_k(\neg m_j))$;
10. $\mathrm{B}_i(\mathrm{B}_i(m_i) \supset r_i)$; 11. $\mathrm{B}_i(\mathrm{B}_j(m_j) \supset r_j)$; 12. $\mathrm{B}_i(\mathrm{B}_k(m_k) \supset r_k)$;
13. $\mathrm{B}_{(i,j)}(\mathrm{B}_i(m_i) \supset r_i)$; 14. $\mathrm{B}_{(i,j)}(\mathrm{B}_j(m_j) \supset r_j)$; 15. $\mathrm{B}_{(i,j)}(\mathrm{B}_k(m_k) \supset r_k)$;
16. $\mathrm{B}_{(i,j)}(\neg r_k))$; 17. $\mathrm{B}_i(\neg r_j))$; 18. $\mathrm{B}_i(\neg m_j \supset \mathrm{B}_i(\neg m_j))$; 19. $\mathrm{B}_i(\neg m_j \supset \mathrm{B}_k(\neg m_j))$;

20. $\mathrm{B}_i(\neg m_k \supset \mathrm{B}_i(\neg m_k))$; 21. $\mathrm{B}_i(\neg m_k \supset \mathrm{B}_j(\neg m_k))$; 22. $\mathrm{B}_{(i,j)}(\neg m_k \supset \mathrm{B}_i(\neg m_k))$; 23. $\mathrm{B}_{(i,j)}(\neg m_k \supset \mathrm{B}_k(\neg m_k))$.

(I): Let the Kripke frame (W, R_1, R_2, R_3) be defined by:

$$\begin{aligned} W &= \{w_1, w_2\} \\ R_i &= R_j = \{(w_1, w_1), (w_2, w_2)\}; \\ R_k &= \{(w_1, w_1), (w_1, w_2), (w_2, w_2)\} \end{aligned} \tag{7.5}$$

The assignment σ is defined by:

$$\sigma(w, p) = \top \text{ iff } w = w_1 \text{ and } p = m_k. \tag{7.6}$$

That is, the only primitive propositional symbol that is ever true is m_k, and it is only true at w_1. For simplicity, we will write $w \models C$ for $(\mathcal{K}, \sigma, w) \models C$.

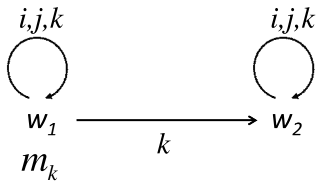


Figure 7.1: Kripke Frame for (I)

This model is depicted in Figure 7.1. An arrow between w and w' with the label s indicates that $(w, w') \in R_s$. Beneath each world, we list the primitive propositional symbols that are true. In this model we exclude belief 3. However, we can verify that every other belief in the list holds true at w_1.

Beliefs 1 and 2 hold since $w_1 \models m_i \vee m_j \vee m_k$. Beliefs 4 through 9, 10, 11, 13, 14, 18 and 19 hold since $w \models \neg m_i$ and $w \models \neg m_j$ for all $w \in W$. Beliefs 12 and 15 hold since $w_2 \models \neg m_k$. Beliefs 16 and 17 hold since $w \models \neg r_k$ and $w \models \neg r_j$ at every $w \in W$. Beliefs 20 to 23 hold since the $w_1 \models m_k$.

It remains only to show that $w_1 \not\models \mathrm{B}_i(m_i)$. This holds since $w \models \neg m_i$ for all $w \in W$.

(II): Let the Kripke frame (W, R_1, R_2, R_3) be defined by:

$$\begin{aligned} W &= \{w_1, w_2, w_3, w_4\} \\ R_i &= \{(w_1, w_2), (w_2, w_2), (w_3, w_3), (w_4, w_4)\}; \\ R_j &= \{(w_1, w_1), (w_2, w_3), (w_3, w_3), (w_4, w_4)\}; \\ R_k &= \{(w_1, w_1), (w_2, w_2), (w_3, w_4), (w_4, w_4)\}. \end{aligned} \tag{7.7}$$

The assignment σ is defined by:

$$\begin{aligned} \sigma(w, m_i) &= \top \text{ iff } w = w_1; \\ \sigma(w, m_j) &= \top \text{ iff } w \in \{w_1, w_2, w_4\}; \\ \sigma(w, m_k) &= \top \text{ iff } w \in \{w_1, w_2, w_3\}; \\ \sigma(w, r_s) &= \top \text{ iff } w = w_2 \text{ and } s = k. \end{aligned} \tag{7.8}$$

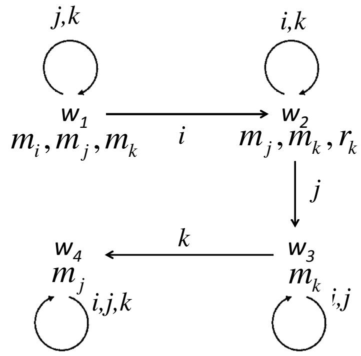


Figure 7.2: Kripke Frame for (II) and (III)

The model is depicted in Figure 7.2. In this model we exclude belief 9. We verify that every other belief in the list holds at w_1.

Beliefs 1, 2, and 3 hold since $w \models m_i \vee m_j \vee m_k$ for all $w \in W$. Belief 4 holds since $w_2 \models \mathrm{B}_j(\neg m_i)$, and belief 5 holds since $w_2 \models \mathrm{B}_k(\neg m_i)$. Belief 6 holds since $w_3 \models \mathrm{B}_j(\neg m_i)$, 7 holds since $w_3 \models \mathrm{B}_k(\neg m_i)$, and 8 holds since $w_3 \models \mathrm{B}_i(\neg m_j)$. Beliefs 10, 11, and 12 hold since $w_2 \models \neg\mathrm{B}_s(m_s)$ for $s = i, j$ and $w_2 \models r_k$. Beliefs 13, 14, 15, and 16 hold since $w_3 \models \neg r_k$ and $w_3 \models \neg\mathrm{B}_s(m_s)$ for $s = i, j, k$. Belief 17 holds since $w \models \neg r_j$ for all $w \in W$. Beliefs 18 and 19 hold since $w_2 \models m_j$. Beliefs 20 to 23 hold since $w_2 \models m_k$ and $w_3 \models m_k$.

It remains only to show that $w_1 \not\models \mathrm{B}_i(m_i)$. This holds since $w_2 \models \neg m_i$.

(III): Let the Kripke frame (W, R_i, R_j, R_k) be the same as in case (II). The assignment σ is:

$$\begin{aligned} \sigma(w, m_i) &= \top \text{ iff } w \in \{w_1, w_4\}; \\ \sigma(w, m_j) &= \top \text{ iff } w \in \{w_1, w_2\}; \\ \sigma(w, m_k) &= \top \text{ iff } w \in \{w_1, w_2, w_3\}; \\ \sigma(w, r_s) &= \top \text{ iff } w = w_2 \text{ and } s = k. \end{aligned} \tag{7.9}$$

We exclude belief 7 in this model. We verify that every other belief in the list holds at w_1.

Beliefs 1, 2, and 3 hold since $w \models m_i \vee m_j \vee m_k$ for all $w \in W$. Belief 4 holds since $w_2 \models \mathrm{B}_j(\neg m_i)$, and 5 holds since $w_2 \models \mathrm{B}_k(\neg m_i)$. Belief 6 holds since $w_3 \models \mathrm{B}_j(\neg m_i)$, and 8 holds since $w_3 \models \mathrm{B}_i(\neg m_j)$. Belief 9 holds since $w_3 \models \mathrm{B}_k(\neg m_j)$. Beliefs 10, 11, and 12 hold for the same reason as in (II). Beliefs 13, 14, 15, and 16 hold since $w_3 \models \neg r_k$, $w_3 \models \neg \mathrm{B}_s(m_s)$ for $s = i, j, k$, and $w_3 \models \neg r_k$. Belief 17 holds since $w \models \neg r_j$ for all $w \in W$. Beliefs 18 and 19 hold since $w_2 \models m_j$. Beliefs 20 to 23 hold since $w_2 \models m_k$ and $w_3 \models m_k$.

It remains only to show that $w_1 \not\models \mathrm{B}_i(m_i)$. This holds since $w_2 \models \neg m_i$.

(IV): Let the Kripke frame (W, R_i, R_j, R_k) be defined by:

$$\begin{aligned} W &= \{w_1, w_2\} \\ R_i &= R_k = \{(w_1, w_1), (w_2, w_2)\}; \\ R_j &= \{(w_1, w_2), (w_2, w_2)\}. \end{aligned} \tag{7.10}$$

The assignment σ is defined by:

$$\begin{aligned} \sigma(w, m_s) &= \top \text{ iff } (w, m_s) \in \{(w_2, m_i), (w_1, m_j)\}; \\ \sigma(w, r_s) &= \top \text{ iff } (w, r_s) = (w_2, r_i). \end{aligned} \tag{7.11}$$

This model is depicted in Figure 7.3.

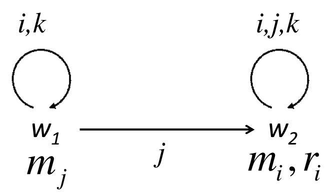


Figure 7.3: Kripke Frame for (IV), (V), (VI)

We exclude belief 4 in this model. We verify that every other belief in the list holds at w_1.

Beliefs 1, 2, and 3 hold since $w \models m_i \vee m_j \vee m_k$ for all $w \in W$. Belief 5 holds since $w_1 \models \mathrm{B}_k(\neg m_i)$. Beliefs 6 and 7 hold since $w_2 \models m_i$. Belief 8 holds since $w_2 \models \mathrm{B}_i(\neg m_j)$, 9 holds since $w_2 \models \mathrm{B}_k(\neg m_j)$. Beliefs 10, 11, and 12 hold since $w_1 \models \neg \mathrm{B}_s(m_s)$ for $s = i, j, k$. Beliefs 13 and 14 and 15 hold since $w_2 \models r_i$, $w_2 \models \neg \mathrm{B}_j(m_j)$, and $w_2 \models \neg \mathrm{B}_k(m_k)$. Belief 16 holds since $w_2 \models \neg r_k$, and 17 holds since $w_2 \models \neg r_j$. Beliefs 18 and 19 hold since

$w_1 \models m_j$. Beliefs 20 to 23 hold since $w \models \neg m_k$ for all $w \in W$.

It remains only to show that $w_1 \not\models \mathrm{B}_i(m_i)$. This holds since $w_1 \models \neg m_i$.

(V) Let the Kripke frame (W, R_1, R_2, R_3) the same one as in (IV). The assignment σ is defined by:

$$\begin{aligned} \sigma(w, m_s) &= \top \text{ iff } s = k; \\ \sigma(w, r_s) &= \top \text{ iff } (w, r_s) = (w_1, r_k). \end{aligned} \tag{7.12}$$

We exclude belief 15, and show that every other belief in the list holds at w_1.

Beliefs 1,2, and 3 hold since $w \models m_i \vee m_j \vee m_k$ for all $w \in W$. Beliefs 4 to 11, 13, and 14 hold since $w \models \neg m_i$ and $w \models \neg m_j$ for all $w \in W$. Belief 16 holds since $w_2 \models \neg r_k$. Beliefs 12 and 17 hold since $w_1 \models r_k$ and $w_1 \models \neg r_j$. Beliefs 18 and 19 hold since $w \models \neg m_j$ for all $w \in W$. Beliefs 20 to 23 hold since $w \models m_k$ for all $w \in W$.

It remains only to show that $w_1 \not\models \mathrm{B}_i(m_i)$. This holds since $w_1 \models \neg m_i$.

(VI) Let the Kripke frame (W, R_1, R_2, R_3) be the same one as is in (IV). The assignment σ is defined by:

$$\begin{aligned} \sigma(w, m_s) &= \top \text{ iff } (w, m_s) \in \{(w_1, m_k), (w_2, m_j)\}; \\ \sigma(w, r_s) &= \top \text{ iff } (w, r_s) \in \{(w_1, r_k), (w_2, r_j)\}. \end{aligned} \tag{7.13}$$

We exclude belief 11 and show that every other belief in the list holds at w_1.

Beliefs 1,2, and 3 hold since $w \models m_i \vee m_j \vee m_k$ for all $w \in W$. Beliefs 4, 5, 6, 7, 10, and 13 hold since $w \models \neg m_i$ for all $w \in W$. Beliefs 8, 9, 14, 15, and 16 hold since $w_2 \models m_j$, $w_2 \models \neg r_k$, $w_2 \models r_j$, and $w_2 \models \neg \mathrm{B}_k(m_k)$. Beliefs 12 and 17 hold since $w_1 \models r_k$ and $w_1 \models \neg r_j$. Beliefs 18 and 19 hold since $w_1 \models \neg m_j$. Beliefs 20 and 21 hold since $w_1 \models m_k$. Finally, beliefs 22 and 23 hold since $w_2 \models \neg m_k$.

It remains only to show that $w_1 \not\models \mathrm{B}_i(m_i)$. This holds since $w_1 \models \neg m_i$. □

Proof of (c): Suppose that $\langle \Gamma_i^1, \Gamma_i^2, \Gamma_i^3 \rangle$ satisfies (a) and (b), but violates (c). Then, by (2.1), $(i, j, k) \notin F_i$. We will show that R5 is violated. By R1 and (b) of this theorem, to show R5 is violated, it suffices to show $\nvdash_{E_i F_i} \Gamma_i^0 \cup \{\mathrm{B}_{(i,j)}(\neg r_k), \mathrm{B}_i(\neg r_j)\} \to \mathrm{B}_i(m_i)$.

Suppose $\vdash_{E_i F_i} \Gamma_i^0 \cup \{\mathrm{B}_{(i,j)}(\neg r_k), \mathrm{B}_i(\neg r_j)\} \to \mathrm{B}_i(m_i)$. Recall $\mathrm{B}_{(i,j,k)}(m_i \vee m_j \vee m_k) = \mathrm{B}_{(i,j)}(\mathrm{B}_k(m_i \vee m_j \vee m_k)) \in \Gamma_i^1$. However, since $(i, j, k) \notin F_i$, the subformula $\mathrm{B}_k(m_i \vee m_j \vee m_k)$ in $\mathrm{B}_{(i,j)}(\mathrm{B}_k(m_i \vee m_j \vee m_k))$ behaves like a propositional variable in the proof of $\Gamma_i^1 \cup \{\mathrm{B}_{(i,j)}(\neg r_k), \mathrm{B}_i(\neg r_j)\} \to \mathrm{B}_i(m_i)$. Hence, we can replace $\mathrm{B}_{(i,j)}(\mathrm{B}_k(m_i \vee m_j \vee m_k))$ by $\mathrm{B}_{(i,j)}(m_i \vee m_j \vee m_k)$ in $\Gamma_i^0 \cup \{\mathrm{B}_{(i,j)}(\neg r_k), \mathrm{B}_i(\neg r_j)\} \to \mathrm{B}_i(m_i)$ without destroying the provability. Since $\mathrm{B}_{(i,j)}(m_i \vee m_j \vee m_k)$ is already in Γ_i^0, the antecedent of the resulting sequent is the belief set $\tilde{\Gamma}_i^3(\mathrm{I})$ of case I of part (b) of this theorem. Now, we would have

$\vdash_{E_iF_i} \tilde{\Gamma}_i^3(\mathrm{I}) \rightarrow \mathrm{B}_i(m_i)$. However, we have shown $\nvdash_{E_iF_i} \tilde{\Gamma}_i^3(\mathrm{I}) \rightarrow \mathrm{B}_i(m_i)$ in part (b) of this theorem. Hence, we conclude that $\nvdash_{E_iF_i} \Gamma_i^0 \cup \{\mathrm{B}_{(i,j)}(\neg r_k),\mathrm{B}_i(\neg r_j)\} \rightarrow \mathrm{B}_i(m_i)$, *a fortiori* R5 is violated. □

References

[1] Aumman, R., (1976), Agreeing to Disagree, *Annals of Statistics* 4, 1236-1239.

[2] Barwise, J., (1981), Scenes and other situations, *Journal of Philosophy*, 78, 369-397.

[3] Binmore, K., (1992), *Fun and Games*, Heath and Company, Lexington.

[4] Bollobas, B., (1953) *Littlewood's Miscellany*, Cambridge: Cambridge University Press.

[5] Geanakoplos, J., and H. Polemarchakis, (1982), We can't disagree forever, *Journal of Economic Theory* 28, 192-200.

[6] Geanakoplos, J. (1992), Common Knowledge, *The Journal of Economic Perspectives*, 6, 53-82.

[7] Gentzen, G.: Untersuchungen ¨uber das logische Schliessen. Mathematische Zeitschrift 39, 176–210, 405–431 (1935). English translation: Investigations into logical deduction. In: Szabo, M. E. (ed.)The Collected Papers of Gerhard Gentzen. Amsterdam: North-Holland 1969.

[8] Gerbrandy, J. and W. Groeneveld, Reasoning about information change, (1997), *Journal of Logic, Language, and Information*, 6, 147-169

[9] Hughes, G. E., and M. J. Cresswell, (1996), *A New Introduction to Modal Logic*, Routledge.

[10] Kaneko, M., (2002), Epistemic Logics and Their Game Theoretic Applications: Introduction, *Economic Theory* 19, 7-62.

[11] Kaneko, M, and T. Nagashima, (1996), Game Logic and its Applications I, *Studia Logica* 57, 325-354.

[12] Kaneko, M., and T. Nagashima, (1997), Game Logic and its Applications II, *Studia Logica* 58, 273-303.

[13] Kaneko, M. and N.Y. Suzuki, (1999), Semantics of Epistemic Logics of Shallow Depths for Game Theory, IPPS. DP. 814, University of Tsukuba.

[14] Kaneko, M. and N.Y. Suzuki, (2002), Bounded Interpersonal Inferences and Decision Making, *Economic Theory* 19, 63-103.

[15] Kaneko, M. and N.Y. Suzuki, (2003), Epistemic Models of Shallow Depths and Decision Making in Games: Horticulture, *Journal of Symbolic Logic* 68, 163-186.

[16] Sato, M., (1977), A study of Kripke-type Models for Some Modal Logics by Gentzen's Sequential Method, *Publications of RIMS* 13: 381-468.

[17] Suzuki, N.-Y., (2012) Semantics for Epistemic Intuisionistic Logics of Shallow Depths, to appear in *Economic Theory* (DOI 10.1007/s00199-012-0707-1).

[18] van Ditmarsch, H., W. van der Hoek, and B. Kooi, (2008), *Dynamic Epistemic Logic*, Springer.

[19] Yasugi, M. and S. Oda, (2002), A Note on the Wise Girls Puzzle, *Economic Theory* 19, 145-156.